1-1-2016

Contextual Insights into the Rett Syndrome Transcriptome

Brian Scott Roosevelt Johnson
University of Pennsylvania, bsj807@gmail.com

Contextual Insights into the Rett Syndrome Transcriptome

Abstract
Mutations in MECP2 are responsible for Rett syndrome (RTT), a severe X-linked neurological disorder characterized by loss of developmental milestones, intellectual disability and motor impairments. However, molecular insight into how these mutations affect the neuronal transcriptiome, disrupt neuronal function and contribute to RTT is impeded by the cellular heterogeneity of the mammalian brain. A comparison between gene expression changes in the striatum, hypothalamus, and cerebellum of MeCP2-null mice revealed that gene expression changes are distinct between different brain regions, which suggests that MeCP2 function should be understood in a cell type-dependent context. To accomplish this task, I generated and phenotypically characterized tagged knock-in mice bearing frequent RTT mutations in Mecp2 (T158M, R106W). Concurrently, I also developed a novel genetic system that allows the tag to be post-translationally modified with biotin in a cell type-specific manner. Altogether, these mice allow for molecular profiling of WT and mutant cortical neurons in the mammalian brain. MeCP2-dependent gene expression changes vary by age and cell type, and the degree of Pol II-mediated transcriptional changes genome-wide correlates with the severity of the RTT mutation. I also detected evidence that supports the post-transcriptional compensation of misregulated long genes, leading to a reinterpretation of prevailing thought within the RTT research field with regards to MeCP2 function. Finally, this genetic strategy circumvents genetic mosaicism associated with female mouse models of RTT and identifies functionally distinct transcriptional changes between neighboring WT and mutant neurons, therefore providing key insights into phenotypic severity between RTT-associated mutation types. By assessing RTT transcriptomes across various neuronal contexts, I propose a novel contextualized paradigm for MeCP2 function in neuronal cell types that not only hypothesizes how MeCP2 dysfunction leads to the cellular deficits that give rise to RTT-like phenotypes, but identifies novel transcriptional and post-transcriptional mechanisms that may be future therapeutic targets for RTT.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Zhaolan Zhou

Keywords
Brain, High-throughput sequencing, MeCP2, Mice, Neurons, Rett syndrome

Subject Categories
Genetics | Molecular Biology | Neuroscience and Neurobiology

CONTEXTUAL INSIGHTS INTO THE RETT SYNDROME TRANSCRIPTOME

Brian S.R. Johnson

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Zhaolan Zhou, Ph.D.

Associate Professor of Genetics

Graduate Group Chairperson

_____

Daniel S. Kessler, Ph.D., Associate Professor of Cell and Developmental Biology

Dissertation Committee

Brian D. Gregory, Ph.D., Assistant Professor of Biology

Hua-Ying Fan, Ph.D., Assistant Professor of Biochemistry and Biophysics

Marisa S. Bartolomei, Ph.D., Professor of Cell and Developmental Biology

Stephen A. Thomas, M.D., Ph.D., Associate Professor of Pharmacology

CONTEXTUAL INSIGHTS INTO THE RETT SYNDROME TRANSCRIPTOME

ACKNOWLEDGEMENTS

I'd first like to thank my advisor, Dr. Zhaolan (Joe) Zhou, who has always seen potential in me from the start and made sure I had the extensive resources and financial support required to explore different techniques and test different hypotheses over the years, regardless of whether such work ended up being fruitful or met a dead end. This experience, coupled with his guidance, has taught me to look forward to the challenges in research as an opportunity to learn and grow regardless of the outcome, making me that much more of an independent scientist.

I would like to acknowledge and thank all past and present members of the Zhou lab for their friendship and support throughout the years, without which I never would have come this far given the broad scope of this project. In no particular order, I'd like to thank the other two members of The Three Amigos that helped to make much of the sequencing work in Chapter 3 possible, Maria Fasolino and Dr. Yingtao (Jerry) Zhao. From 4am experiments to in-depth discussions on how to analyze data, you both have had the largest influence on the success of this project. I'd like to thank Dr. Janine Lamonica and my Korean big sister, Dr. Deborah Kwon, who have been my mentors that provided a lot of emotional support and advice during my last few months of graduate school. To the current crop of graduate students, Barbara Terzic, Sarah Welsh, Alice Dallstream, and Sheng Tang, you guys have been instrumental in both helping me to form the ideas that have gone into the discussion for this thesis, and making sure I can always have the occasional fun night out. A special shout out to my undergraduate mentee, Daniel Bu, not only for his work in quantifying some of my IHC experiments, but for also helping to keeping me young and up to date on pop culture trends. To former lab members, I'd like to return the thanks to Dr. Kathleen Wood, who has been my laboratory baymate throughout the years and who knows that the struggle is real. I'd also like to thank the original Zhou lab crew, Megan Allen, and Drs. Darren Goffin, Judy Wang, and Maria Amorim for making me feel welcome when I first joined the Zhou lab all those years ago, and providing so much social, intellectual, and emotional support as I initiated and carried out this project. You truly made the lab feel like a second home

to me. Finally, I'd especially like to thank Yolanda Cui for her hard work in keeping our lab running smoothly and maintaining our mouse colonies, especially my very large mouse colony.

I'd also like to acknowledge those outside the Zhou lab, including my neighbors in Dr. Doug Epstein's laboratory, which always felt like a lab away from… well, lab. Thank you Alex Rohacek, Staci Rakowiecki, Victor Muthu for your friendship over the years. A special thanks goes to Yao Yao, who was always there to listen when I needed someone to talk to about research or otherwise, and who is always willing to cheer me on. Another special thanks goes to Tanya Corman. We've been through a lot throughout grad school and beyond, from prelims, to birthdays, to softball and the 'hood… I can honestly say my time in grad school and my social life in Philly have been uniquely shaped by your valuable friendship, humor, timing (one of them is European…) and support throughout the years. I would not have made so much intellectual and personal growth in Philly had you not overheard my conversation at Steve's party or ended up just a few bays down from me. I'd also like to thank the people associated with my first laboratory experience at Penn, in particularly Dr. Aaron Gitler, who first recruited me to Penn as a technician and then helped me apply to graduate school. I can't imagine the direction my life would have gone in without his mentorship and advice throughout the years.

I'd like to thank members of my thesis committee, Drs. Brian Gregory, Hua-Ying Fan, Steve Thomas, and Marisa Bartolomei, who always balanced out my critical nature with encouraging words and great insights that ultimately made my research even better than I would have imagined it to be. Another special thanks to Marisa, who also had to put up with me and my neurotic way of thinking during the times when I rotated in her lab, when she served on my preliminary examination committee, and when she mentored me as a student on the CAMB T32 training grant together with Dr. Richard Schultz. I'd also like to thank Drs. Thomas Jongens, Doug Epstein, and Meera Sundaram for their mentorship and support as current and former heads of the Genetics and Gene Regulation graduate group at Penn, as well as the head of CAMB, Dr. Dan Kessler, and the CAMB coordinators Megan Schofer, Anna Kline, and Kathy O'Conner-

Cooley for their support in running in graduate program. I'd also like to acknowledge William DeMuth and Hank Pletcher at the Flow Cytometry and Cell Sorting facility, Jonathan Schug and associates at the Next-Generation Sequencing Core facility, the Cell Center Stockroom, all staff members of the University Laboratory Animal Resources, and the Penn Genetics department for helping to make this project possible.

Finally, I want to thank my friends for all of their support over the years, including fellow classmates such as Dustin Hancks, Ian Silverman, and Alex Brown among countless others that have helped me to move forward when times were tough. Special shout-out to Kate Palozola and Justin Becker with whom I had so much fun at a Santa Fe Keystone conference, it ruined future conferences for all of us, which we agree won't even be half as much fun. I'd like to thank my close friends who have provided me support over the years, including my former roommate Aaron Hann, who is always great for California excursions and being featured in a movie that streams on Netflix ("The Circle" – go watch it), and my mentor Joseph Chimenti, who is always fun to be around and helps keep my head on straight. Thank you all for your encouragement and guidance over the years.

ABSTRACT

CONTEXTUAL INSIGHTS INTO THE RETT SYNDROME TRANSCRIPTOME

Brian S.R. Johnson

Zhaolan Zhou, Ph.D.

Mutations in *MECP2* are responsible for Rett syndrome (RTT), a severe X-linked neurological disorder characterized by loss of developmental milestones, intellectual disability and motor impairments. However, molecular insight into how these mutations affect the neuronal transcriptiome, disrupt neuronal function and contribute to RTT is impeded by the cellular heterogeneity of the mammalian brain. A comparison between gene expression changes in the striatum, hypothalamus, and cerebellum of *MeCP2*-null mice revealed that gene expression changes are distinct between different brain regions, which suggests that MeCP2 function should be understood in a cell type-dependent context. To accomplish this task, I generated and phenotypically characterized tagged knock-in mice bearing frequent RTT mutations in *Mecp2* (T158M, R106W). Concurrently, I also developed a novel genetic system that allows the tag to be post-translationally modified with biotin in a cell type-specific manner. Altogether, these mice allow for molecular profiling of WT and mutant cortical neurons in the mammalian brain. MeCP2-dependent gene expression changes vary by age and cell type, and the degree of Pol II-mediated transcriptional changes genome-wide correlates with the severity of the RTT mutation. I also detected evidence that supports the post-transcriptional compensation of misregulated long genes, leading to a reinterpretation of prevailing thought within the RTT research field with regards to MeCP2 function. Finally, this genetic strategy circumvents genetic mosaicism associated with female mouse models of RTT and identifies functionally distinct transcriptional changes between neighboring WT and mutant neurons, therefore providing key insights into phenotypic severity between RTT-associated mutation types. By assessing RTT transcriptomes across various neuronal contexts, I propose a novel contextualized paradigm for MeCP2 function

in neuronal cell types that not only hypothesizes how MeCP2 dysfunction leads to the cellular deficits that give rise to RTT-like phenotypes, but identifies novel transcriptional and post-transcriptional mechanisms that may be future therapeutic targets for RTT.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: Introduction

1 General Introduction

1.1 THE *MECP2* GENE AND PROTEIN

1.1.1 *MECP2* Gene and Protein Structure and Function

Methyl-CpG binding protein 2 (MeCP2) was first detected in a biochemical screen to identify and clone novel protein factors from crude rodent nuclear extracts that were capable of binding to a single symmetrically methylated CpG (Lewis et al., 1992). In this initial study, MeCP2 was described as an 84kDa protein that demonstrated robust binding to methylated CpG, but showed weak or undetectable binding for hemi-methylated or non-methylated CpGs, respectively. Expression cloning was used to identify the ORF from a rat brain cDNA library. Strangely, this ORF was predicted to encode a protein of 53kDa, suggesting that MeCP2 exhibits lower mobility on an SDS-PAGE gel than expected. Immunofluorescent staining in cultured rat cells using an antibody raised against MeCP2 revealed that this protein was likely chromatin bound in the nucleus, which was confirmed through its release from mouse nuclei following enzymatic digest of DNA. In addition to being uniformly distributed throughout the mouse nucleus, MeCP2 also demonstrated strong punctate nuclear staining due to association with pericentromeric chromatin, which suggested that MeCP2 is distributed over euchromatic and heterochromatic regions. These nuclear puncta, which was estimated to contain over 40% of all methyl-CpGs in the mouse nucleus, are enriched for satellite repeat units, each of which are AT-rich and contain at least seven methylated CpGs per unit. Thus MeCP2 was initially described as a chromatin associated nuclear protein that binds to methyl-CpG-rich containing DNA.

MeCP2 is a founding member of the methyl-CpG binding domain (MBD) family of proteins that have binding affinity for methylated DNA (Lewis et al., 1992). In addition to the MBD, which principally dictates MeCP2 interactions with symmetrically methylated CpGs, MeCP2 also

contains a transcriptional repression domain (TRD) which mediates protein-protein interactions with various histone deacetylase (HDAC) co-repressor complexes, including the NCoR-SMRT and Sin3A-HDAC1&2 comples (Ebert et al., 2013; Nan et al., 1996, 1997, 1998). Hence MeCP2 is traditionally known to bind to methylated DNA and mediate transcriptional repression *in vivo*. MeCP2 also contains a Nuclear Localization Sequence (NLS) within the interior of the TRD and mediates its localization to the nucleus (Lewis et al., 1992). In addition to the two principal domains (MBD and TRD) that are critical for MeCP2 function, MeCP2 also has a long C-terminal Domain (CTD) that, although unstructured, may be important for MeCP2 binding in a methylation-independent manner (Heckman et al., 2014; Nikitina et al., 2007a). The CTD may also play a role in modulating chromatin structure, as several AT-hook domains, which closely resemble those found in the family of high-mobility group proteins, are spaced throughout the C-terminal tail and are required for proper binding of MeCP2 to chromatin *in vivo* (Baker et al., 2013).

Aside from the well conserved and highly structured MBD domain, more than 60% of MeCP2 is intrinsically disordered and does not adopt strong secondary protein structure, although the protein becomes more structured upon binding to methylated DNA (Adams et al., 2007). This unusual property is thought to underlie the multifunctional roles associated with MeCP2 as a chromatin-bound architectural protein that can modulate transcription. Like most unstructured proteins, MeCP2 has thus been shown to interact with multiple protein partners that help define its function. In addition to the NCoR-SMRT and Sin3a complexes, MeCP2 has also been shown to interact with additional co-repressors, including cSki, and CoREST (Abuhatzira et al., 2007; Kokura et al., 2001), transcriptional activators such as CREB (Chahrour et al., 2008), and RNA-binding proteins that may influence RNA splicing, such as YB1 and TDP-43 (Sephton et al., 2011; Young et al., 2005). MeCP2 is also known for interacting with proteins that regulate epigenetic alterations to chromatin structure, including, but not limited to, ATRX, Dnmt1, Tet1, and HP1 (Cartron et al., 2013; Chatr-Aryamontri et al., 2013; Kernohan et al., 2014; Nan et al., 2007).

The human *MECP2* gene is mapped to the Xq28 gene locus, and is thus subject to X-inactivation in females. The gene encompasses 152kb of genomic sequence in humans and 123kb of genomic sequence in mice, and is comprised of four coding exons. While exons 1 and 2 are alternatively spliced (described below) and encode the N-terminal domain of MeCP2, exons 3 and 4 encode both halves of the MBD, and exon 4 encodes the NLS, TRD, and remainder of the C-terminal domain, as well as the 3'-UTR (described below). Transcriptional regulation of MeCP2 is likely to be extensive, as the second intron of MeCP2 (between exons 2 and 3) is atypically long at 60kb and contains several regions with high evolutionary conservation (Singh et al., 2008) and likely contains multiple regulatory elements. A recent study performing a comparative sequence analysis of the *MECP2*-gene in both human and mice found 170 conserved elements >50bp in length, five of which are found with *MECP2* exon sequences, and 109 of which are found in untranslated regions and intron sequences (Reichwald et al., 2014).

Exon 1 and exon 2 of the *MECP2* gene each contain an ORF and are subject to alternative splicing, producing two translated protein products that, aside from unique amino termini that differ in size by 13 amino acids, remain 96% similar in primary sequence. The MeCP2-e1 isoform (which was previously known as MeCP2B due to its order of discovery, or MeCP2α), which is derived from exons 1,3, and 4, is 498 amino acids long and is more abundantly expressed in neurons than glia throughout the mouse brain, although it's abundance varies between individual brain regions (Dragich et al., 2007; Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004; Zachariah et al., 2012). In contrast, the MeCP2-e2 isoform (previously known as MECP2A or MeCP2β) derived from exons 2-4 is 486 amino acids long, with sequences in exon 1 contributing to the 5'-UTR that are required for translation of MeCP2-e2 (Saxena et al., 2006). Although the overall abundance of MeCP2-e2 in the mouse brain is significantly lower (~10X less abundant) relative to MeCP2-e1, MeCP2-e2 demonstrates a more restricted spatiotemporal pattern of expression, being widely expressed during early postnatal development before being restricted to the dorsal thalamus and cortical laver V of the brain (Dragich et al., 2007). Despite the differences in abundance and patterns of expression, the high similarity in C-

terminal sequence suggests that both isoforms appear to be functionally equivalent, and the transgenic expression of each isoform in mice demonstrates that one isoform can functionally compensate for the other (Kerr et al., 2012). However, the lack of pathogenic mutations in exon 2 of the *MECP2* gene, in contrast to several that have been identified in exon 1, suggests that MeCP2-e2 is likely indispensible for function in contrast to MeCP2-e1 (Gianakopoulos et al., 2012; Itoh et al., 2012; Saunders et al., 2009; Saxena et al., 2006). As a result, the remainder of this section will only consider MeCP2 in the context of the MeCP2-e1 isoform.

MeCP2 is expressed in multiple tissues and organs of the human and mouse body to a variable degree. MeCP2 protein expression is highest in the brain, lung, and spleen, lower in the heart and kidney, and barely detectable in liver, stomach, and small intestine (Shahbazian et al., 2002b). However, MeCP2 RNA transcripts is detected in all human and mouse tissues, which suggests that this protein is significantly regulated at a post-transcriptional level (Shahbazian et al., 2002b). Post-transcriptional regulation likely arises from it's 3'-UTR in exon 4, which represents one of the longest 3'-UTRs in the human genome and is subject to alternative polyadenylation usage, resulting in mRNA transcripts 1.9kb, 5kb, 7.5kb and 10.1kb in size (Coy et al., 1999; Reichwald et al., 2014; Singh et al., 2008). These transcripts exhibit tissue-specific expression patterns but have almost identical half lives, and the shortest and longest transcripts contain AU-rich elements which are known to be associated with reduced RNA transcript stability (Reichwald et al., 2014). The shortest transcript is detectable in heart, skeletal, muscle, and spleen tissues, whereas the longest transcript is detectable in skeletal muscle, kidney, pancreas, and brain (Reichwald et al., 2014). The longest 3'-UTR (10.1kb), exhibits between 52%-74% sequence homology between human and mice (depending on the study), and exhibits low ubiquitous expression in early organogenesis but higher expression in the mature brain (Coy et al., 1999; Reichwald et al., 2014). Furthermore, changes in polyadenylation usage over age result in the postnatal decrease, than increase in usage of the long 3'-UTR, which negatively correlates with postnatal changes in MeCP2 expression as mice age (Balmer et al., 2002; Pelka et al., 2005; Shahbazian et al., 2002b). The evolutionary conservation of the long 3'-UTR is limited to blocks of

primary sequences that are highly conserved across mice, humans kangaroo, rat, hamster, macaque, and chimpanzee sequences and suggests the potential for post-transcriptional regulation by miRNAs (Coy et al., 1999; Klein et al., 2007). One brain-enriched and activity-dependent miRNA, miR-132, has been shown to recognize this long 3'-UTR and downregulate MeCP2 expression in response to neuronal depolarization in primary cortical neurons (Klein et al., 2007). Interestingly, this study found that miR-132 expression is induced through activity-dependent expression of *Bdnf*, a well-known MeCP2 target gene, which suggests that MeCP2 may indirectly downregulate its expression in the brain in response to neuronal activity as a means of homeostatic control. Similar homeostatic loops have been identified for miRNA-212 in mediating cocaine responses in the striatum, miR-7b during postnatal maturation of *Mecp2* expression in neurons, among others (Chen et al., 2014; Im et al., 2010), suggesting complex neurodevelopmental regulation of MeCP2 expression that may change with age.

MeCP2 protein is highest in post-mitotic neurons of the mammalian brain. Neuronal MeCP2 is relatively low a birth, but dramatically increases with age as neurons mature and undergo synaptogenesis (Alvarez-Saavedra et al., 2007; Chen et al., 2015; Kishi and Macklis, 2010; Shahbazian et al., 2002b; Skene et al., 2010). The rapid increase in MeCP2 expression levels is particularly steep within the mouse neocortex and hippocampus, which undergo synaptogenesis within the first two weeks of murine post-natal life, whereas the cerebellum displays a delayed increase in expression after four weeks of age (Mullaney et al., 2004). Biochemical quantification and comparison of MeCP2 abundance in the adult mouse revealed that each post-mitotic neuronal nucleus contains 16 million molecules of MeCP2, in contrast to 2 million molecules of MeCP2 in glial nuclei (8 times less abundant) and 500,000 molecules of MeCP2 in liver hepatocyte nuclei (32 times less abundant) (Skene et al., 2010). This indicates that post-mitotic neurons contain 1 molecule of MeCP2 for every 2 nucleosomes, enough to potentially saturate all methyl-CpGs and coat the neuronal genome (Skene et al., 2010). Subsequent studies have shown that the high expression of MeCP2 protein in the brain is carefully balanced, as too much (Collins et al., 2004; Friez et al., 2006; del Gaudio et al., 2006;

Liu et al., 2016; Meins et al., 2005) or too little (Samaco et al., 2008) MeCP2 protein results in neurological phenotypes. Thus the high expression of MeCP2 in the brain, which is tightly regulated by a complex array of transcriptional and post-transcriptional mechanisms that differ between brain regions, argues for MeCP2 playing an important role in CNS function.

1.2 THE CLINICAL RELEVANCE OF *MECP2*

1.2.1 Clinical Features of Rett Syndrome

Rett syndrome (RTT) is named after the Austrian pediatrician Dr. Andreas Rett, who first described a curious condition among 22 patients that resulted in severe brain atrophy and relentless hand-wringing behavior, which he, at the time, attributed to hyperammonemia (Rett, 1966). However, this syndrome went overlooked until 1983, when Swedish neurologist Dr. Bengt Hagberg reported 33 clinical cases across three different countries, all of which exhibited a uniform and strikingly progressive X-linked encephalopathy reminiscent of what had been described 17 years prior (Hagberg et al., 1983). The specific clinical features that were described, combined with the identification of additional patients worldwide, further established RTT as a clinical syndrome. However, the molecular etiology of the disorder was unknown until 1999, when pediatric neurologist Dr. Huda Zoghbi and her lab discovered both rare familial and common sporadic *de novo* mutations that associate with RTT, which were quickly followed by the discovery of additional mutations and genomic aberrations at this gene locus (Amir et al., 1999). Since then, research interest has exploded in order to discover the exact nature of MeCP2 function within the CNS, and how molecular dysfunction of this protein leads to neuronal deficits and human disease.

RTT is a progressive X-linked neurological disorder that represents one of the most common causes of mental disability among young girls, occurring in 1 out of every 10,000 female births. Individuals with RTT exhibit wide variability in the rate of disease progression and severity, and can broadly be classified into typical (stereotyped RTT progression) and atypical (RTT variant progression) RTT cases (Cuddapah et al., 2014; Neul et al., 2010). Individuals with typical

RTT appear asymptomatic during early childhood development. During the initial phase of RTT onset, which occurs ~6-18 months of age, many of these patients begin to exhibit developmental delays, including a failure to meet developmental milestones, a deceleration in the rate of head and body growth appropriate for their age (microcephaly), and reduced muscle tone (hypotonia) (Chahrour and Zoghbi, 2007; Hagberg et al., 1983; Tarquinio et al., 2012). This is followed by a phase of rapid developmental regression in which these patients lose acquired developmental skills over a period of weeks to a year, including purposeful hand use, spoken language, and the ability to walk. Patients during this period also experience a temporary indifference to the surrounding environment and unresponsiveness to social cues, which caused RTT to be characterized as an autistic-spectrum disorder up until it DSM-V was released in 2013. Many of the characteristic clinical features associated with RTT arise during this regression phase, including the “hand-wringing” hand stereotypies, impaired motor coordination and ataxia, and intellectual disability. RTT patients also display severe autonomic dysfunction that is characterized by irregular respiratory behaviors such as breath holding and hyperventilation during waking hours, forced expulsion of air and saliva, and sleep apneas (Chahrour and Zoghbi, 2007; Ramirez et al., 2013). After regression, many patients experience a pseudostationary stage in which symptoms begin to stabilize; some individuals may display subtle improvements and partially regain previously lost skills (Lombardi et al., 2015; Neul et al., 2010), which suggests that RTT symptoms may be reversible. Notably, many RTT patients are socially aware of their environment, which is typified by the appearance of an intense eye-gaze that these patients display as a method of communication (Kaufmann et al., 2012). A large proportion of RTT patients also exhibit seizures that begin during early childhood and range from easily controllable to severe epilepsy, although this symptom may decline in frequency during adolescence (Jian et al., 2006). From adolescence into adulthood, many patients enter a stationary phase characterized by scoliosis, increase autonomic dysfunction, and increasing loss of mobility associated with Parkinsonian-like features (Chahrour and Zoghbi, 2007).

Atypical RTT patients represent a small proportion of the total RTT patient population (14% from a 2014 study) can be described as lying on the phenotypic extremes of typical RTT. There are three associated variants: atypical RTT with preserved speech, early onset seizures, or congenital variants. Atypical RTT with preserved speech is the least severe variant, which includes mildly affected individuals who can walk, talk and draw; in contrast, atypical individuals with the congenital variant never acquire the ability to speak and have difficulty sitting (Ariani et al., 2008; Artuso et al., 2010; Cuddapah et al., 2014; Renieri et al., 2009; Zappella et al., 2003). What remains notable about atypical RTT patients is that they exhibit RTT clinical symptoms despite a failure to detect *MECP2* mutations among some individuals. In recent years, however, additional research has revealed that the early seizure RTT variant is genetically linked to Cyclin-dependent kinase-like 5 (*CDKL5*), a protein-kinase in the CNS that may be involved in synaptic plasticity, and is now recognized as the clinically distinct *CDKL5* disorder (Artuso et al., 2010). Similarly, the congenital RTT variant is genetically linked to *Forkhead box G1* (*FOXG1*), a transcriptional regulator, and is now recognized as the clinically distinct *FOXG1* syndrome (Ariani et al., 2008).

1.2.2 MeCP2 and the Genetics of Rett Syndrome

More than 95% of RTT clinical cases can be linked to mutations in the *MECP2* gene, which to date total 888 unique nucleotide changes (missense, nonsense, deletions, insertions, duplications) across the *MECP2* gene. However, 58% of RTT cases have 1 of 8 frequent mutations, which all occur at CpG hotspots and likely arise from the deamination of methylated cytosines. MeCP2 R106W, R133C, and T158M are missense mutations located in the MBD of MeCP2 and have been shown to disrupt MeCP2 binding to methylated DNA (Ballestar et al., 2000; Brown et al., 2015; Ghosh et al., 2008; Ho et al., 2008) by altering the hydrophobic environment of the ASX-ST motif within the MBD (Ghosh et al., 2008; Ho et al., 2008). R306C is another missense mutation that resides within the TRD and has been shown to disrupt MeCP2's interaction with the Nuclear Receptor Corepressor Complex (NCoR, see section 1.4.3 for more detailed information) (Brown et al., 2015; Ebert et al., 2013; Lyst et al., 2013). The remaining four

mutations are nonsense mutations. R168X is a nonsense mutation that occurs 3' to the MBD. R255X and R270X are nonsense mutations that overlap with the NLS sequence (amino acids 255-271), although R270X can still be translocated to the nucleus, while R294X disrupts the TRD.

The majority of *MECP2* mutations occurs *de novo* and arises spontaneously in the paternal germ line. Because *MECP2* is an X-linked gene, most mutations will typically be inherited by female offspring that are heterozygous for the mutation. Random X-inactivation (XCI) in females will cause female RTT patients to display variable clinical features that are dependent on the percentage of somatic cells expressing the wild-type or mutant *MECP2* allele, in addition to the severity of the mutation. In some cases, XCI can be non-random, which skews expression towards the wild-type or mutant allele and leads to much milder (akin to a mild learning disability) or more severe clinical symptoms, respectively (Weaving et al., 2003). Familial cases of RTT, as well as RTT cases among males, are incredibly rare. However, owing to non-random XCI that skews towards the wild-type allele in germline cells, many mosaic females can become carriers and pass their *MECP2* mutant allele to their offspring. Hemizygous males that inherit *MECP2* mutations are severely affected, and either display severe intellectual disability, present with neonatal encephalopathy and rarely survive beyond two years of age, or die *in utero* owing to the nature and severity of the mutation (Meloni et al., 2000; Neul et al., 2010; Zeev et al., 2002, 2002)

XCI confounds many clinical studies that have tried to establish clear genotype-phenotype correlations between mutation type/severity and specific RTT clinical outcomes, especially since assessments of XCI is limited to circulating leukocytes from blood draws, which may not represent XCI in the tissue of interest (Archer et al., 2007; Cuddapah et al., 2014; Temudo et al., 2011; Wu et al., 2014). This is compounded by the difficulty in studying rare males bearing *MECP2* mutant alleles that are so severely affected. Nonetheless, broad associations have emerged that find that the nature of the *MECP2* mutation affects RTT severity (Bebbington et al., 2008; Charman et al., 2005; Cuddapah et al., 2014; Young et al., 2011). Most missense mutations occur in two broad clusters that overlap with the MBD and the TRD of MeCP2,

highlighting the functional and clinical importance of these two domains. Several studies have corroborated that mutations which are predicted to disrupt MeCP2's ability to bind methylated DNA are associated with higher clinical severity scores and more severe symptoms that are diagnosed at earlier ages, on par with deletions, insertions, and splice site mutations that are damaging for protein function. In contrast, mutations less likely to disrupt binding to methylated DNA are associated with lower clinical severity scores. Interestingly, the T158M mutation exhibits the most variability between studies and is classified as a severe mutation in some studies and an intermediate mutation in others (Bebbington et al., 2008; Charman et al., 2005; Cuddapah et al., 2014; Young et al., 2011), which might be attributed to the degree of XCI (Archer et al., 2007).

1.3 MODELING MECP2 MUTATIONS IN RETT SYNDROME

1.3.1 Mouse Models of Rett syndrome

Germline and Conditional Deletion of *Mecp2*

The first mouse models of MeCP2 were generated by either deleting exon 3, which generated a truncated protein (Chen et al., 2001), or exon 3 and 4, which completely abolished a detectable protein product (Guy et al., 2001). Notably, in both studies, mice have lost the MBD domain, resulting in a failure of MeCP2 binding to methylated DNA *in vivo*. *Mecp2*-null male mice display striking physiological changes that began around 4-6 weeks of age, including body tremors, gain or loss of body weight depending on the genetic background, abnormal gait, hindlimb clasping, and premature lethality around 2 months of age (Chen et al., 2001; Guy et al., 2001). In addition to the apparently normal early development prior to symptom onset, these *Mecp2*-null mice also displayed reduced brain size and respiratory abnormalities, which are prominent features that are observed among the RTT patient population (Neul et al., 2010). The reduced brain size is associated with increase neuronal cell packing without gross changes in brain anatomy and morphology, suggesting that RTT is not a product of neurodegeneration (Chen et al., 2001; Guy et al., 2001). In contrast, *Mecp2*-null female mice display a delayed onset of RTT-like symptoms (~3-4 months of age), but by 12 months of age display similar features as

male mice, including tremor, abnormal gain, hypoactivity, hindlimb clasping, and respiratory abnormalities (Chen et al., 2001; Guy et al., 2001).

Given that germ line loss of MeCP2 in mice results in the appearance of neurological RTT-like phenotypes, a key question in the field has been whether these phenotypes are attributed to deficits in neuronal function. A seminal experiment that addressed this question involved the conditional removal of MeCP2 from the whole brain using a Nestin-Cre driver, which depletes MeCP2 expression from most post-mitotic neurons and glia in the CNS. Conditional deletion of MeCP2 from the brain was sufficient to recapitulate the phenotypes observed in germ line *Mecp2*-null mice (Chen et al., 2001; Guy et al., 2001), confirming that all RTT phenotypes are primarily neurological in origin, although mutant glia may also contribute to RTT pathogenesis in ways that are not fully understood (Lioy et al., 2011).

Given that global loss of MeCP2 is associated with multiple cognitive, motor, and physiological symptoms, there has been great interest in dissecting the RTT phenotypes even further and determining whether particular neuronal cell types are responsible for neurological deficits. Although specific cell populations correlate with specific phenotypes upon loss of MeCP2 (see Table 1.2), most subpopulations of neuronal cells are not sufficient to recapitulate all aspects of RTT phenotypes and are rather variable. For example, the loss of MeCP2 function from all GABAergic neurons in the CNS is sufficient to recapitulate all RTT-like phenotypes because GABAergic neuronal function is important for ensuring that every neuronal circuit in the brain is functioning in a coordinated fashion (Chao et al., 2010). In contrast, the loss of MeCP2 function from all forebrain GABAergic neurons results in milder RTT-like phenotypes, likely because neuronal circuits in the hindbrain are not being affected by lost of GABAergic inhibitory signaling, which would otherwise impact the autonomic functions of that brain region that largely contribute to RTT phenotypes (Chao et al., 2010; Ward et al., 2011). In contrast, the loss of MeCP2 function from all forebrain excitatory neurons only results in mild abnormal cognitive behaviors, characteristic of those neurons mediating higher order cognitive function (Gemelli et al., 2006).

Collectively, these studies have found that the variability in RTT phenotypic impairments correlate with the brain region or circuit that is being adversely affected. However, it remains unclear whether cell type-specific impairments in neuronal function correlate with cell type-specific gene expression changes in the absence of MeCP2.

Postnatal Deletion and Re-activation of *Mecp2*

Postnatal deletion of the *Mecp2* gene in late juvenile or adult male and female mice across three independent studies all result in the same rate of onset of RTT-like symptoms and premature lethality seen with germ line *Mecp2*-null male and female mice regardless of the age of inactivation (Cheval et al., 2012; McGraw et al., 2011; Nguyen et al., 2012). This includes higher neuronal cell packing density and decreased branch complexity and spine density of neuronal dendrites which altogether contribute to reduced brain weight (Nguyen et al., 2012) and are hallmarks of RTT symptoms in human patients. In one of these studies, hippocampal neurons also display reductions in the protein expression of synaptic genes such as CaMKII α/β, AMPA and NMDA receptors, and the synaptic vesicle proteins Vglut and Synapsin, which is indicative of impaired synaptic structural plasticity and overall neuronal function (Nguyen et al., 2012). Notably, the decreases in synaptic proteins were not reflected among unchanged mRNA levels; although MeCP2 may be involved in the post-transcriptional regulation of these genes, these observations are entirely consistent with studies that found that MeCP2 depletion results in reduced transcription of ribosomal RNAs, leading to global deficits in protein translational rates (Li et al., 2013a; Yazdani et al., 2012). These observations suggest that MeCP2 transcriptional regulation may be actively and continuously required to maintain normal neuronal functions.

Inversely, several studies have now demonstrated that RTT-like phenotypes can be genetically reversed in *Mecp2*-null male mice. By removing a floxed Stop cassette from the *Mecp2* gene locus (thus reactivating expression from the endogenous mouse gene) or from a floxed Stop-*Mecp2* transgene in combination with various Cre lines (Nestin-Cre: all neurons and glia starting at embryonic day 10; Tau-Cre: all post-mitotic neurons starting at embryonic day 9.5;

CaMKII-Cre: postnatal forebrain neurons starting at post-natal day 0-15 or 15-30, depending on the Cre line), these studies have shown that post-natal restoration of MeCP2 expression in previously *Mecp2*-null post-mitotic neurons of male and female mice is sufficient to restore brain size and body weight (Giacometti et al., 2007; Guy et al., 2007). In female mice, the post-natal restoration of MeCP2 expression also rescues electrophysiological deficits in long-term potentiation and synaptic plasticity (Guy et al., 2007). Lifespan is also noticeably extended when *Mecp2* expression is restored in all post-mitotic neurons in the CNS (>8 months survival) (Giacometti et al., 2007). However, lifespan for *CamKII*-Cre mice was only extended by ~4 weeks, suggesting that proper MeCP2 expression in hindbrain post-mitoic neurons, such as HoxB1+ neurons in the brainstem, likely mediate the extensions in lifespan (Giacometti et al., 2007; Ward et al., 2011).

Interestingly, it has also been found that although *Mecp2* expression is low in glia and is not necessary for initiating RTT phenotypes, the conditional restoration of MeCP2 expression in glial cells is sufficient to significantly improve locomotion and anxiety levels, restore respiratory abnormalities, and prolong the lifespan of *Mecp2*-null animals, along with restoring normal dendritic morphology and increasing VGLUT1 protein expression (Lioy et al., 2011). This study suggests that non-cell autonomous effects may also play a role in RTT pathogenesis, and that the secreted trophic support provided by glia in the CNS may be viable targets for RTT therapeutics.

Altogether, genetic studies using various mouse models confirm that MeCP2 is involved in the active and dynamic maintenance of post-mitotic neuronal function, and that the absence of MeCP2 at earlier stages is not a pre-requisite for manifesting RTT-like phenotypes, nor a barrier towards the development of therapeutic interventions.

Mecp2 RTT-Associated Point Mutations

Given the difficulty in performing genotype-phenotype correlations among clinically variable human RTT female patients (Cuddapah et al., 2014), the development of mouse models

bearing RTT-associated missense and non-sense mutations has expanded our molecular understanding of MeCP2 function and the impact of individual pathogenic mutations (Figure 1.2, Table 1.1). These mouse models have been shown to faithfully model the degree of phenotypic severity observed in human RTT patients, thus providing key insights into the molecular aspects of RTT.

Mutations in the MBD of MeCP2 are known to variably impact RTT patients (see Figure 1.2), which are likely mediated by differences in MeCP2 binding affinity to methylated DNA. *In vitro* studies have shown that MeCP2 R133C binding to methylated CpGs is indistinguishable from wildtype MeCP2, whereas MeCP2 T158M retains partial affinity, and MeCP2 R106W completely abolishes interactions between the MBD and methylated CpGs (Ballestar et al., 2000; Brown et al., 2015; Ghosh et al., 2008; Ho et al., 2008). To date, MeCP2 R133C, T158M, and T158A knock-in mutations have been modeled in mice (Brown et al., 2015; Goffin et al., 2012). MeCP2 R133C is noticeably less severe when compared to T158M/A mutations, best characterized by the differences in symptom onset (R133C: 42 weeks, T158M: 13 weeks, T158A: 16 weeks). The differences in lifespan correlate with the expected impairments in binding to methylated DNA, suggesting that the severity of mutations in the MBD is dictated by the degree of impairment to methylated DNA binding. This is further supported by the generation of MeCP2 R111G mutant mice, which appear phenotypically similar to *Mecp2*-null mice in terms of symptoms and lifespan (12 weeks), consistent with the complete abolishment of binding to methylated DNA (Free et al., 2001; Kudo et al., 2003).

Mice bearing RTT-associated point mutations in the TRD of *Mecp2* confirm that the clinical variability associated with this domain is dependent on how much of the TRD function is conserved (see Table 1.1). Mice with R168X and R255X are predicted to truncate MeCP2 prior to its NLS, and thus should not be present in the nucleus. However, studies were unable to detect either full length or partial MeCP2 protein products (Lawson-Yuen et al., 2007; Pitcher et al., 2015; Wegener et al., 2014). As a result, both R168X and R255X male mice appear

phenotypically similar to *Mecp2*-null male mice and display premature lethality, respiratory abnormalities, and impairments in motor function, effectively making these mutations null alleles that are consistent with the clinical severity observed in human RTT patients.

In contrast, among three independent lines of mice bearing R306C mutations (Brown et al., 2015; Heckman et al., 2014; Lyst et al., 2013), retention of the full TRD results in relatively mild disease severity and increased lifespan when compared to *Mecp2*-null mice, consistent with the mild clinical severity of R306C-bearing human patients. However, R306C mice continue to display deficits in locomotion, motor coordination, and learning and memory, similar to other *Mecp2*-mutant mice (Brown et al., 2015; Heckman et al., 2014; Lyst et al., 2013). Interestingly, two of the three R306C lines display a median lifespan of ~18 weeks (Heckman et al., 2014; Lyst et al., 2013), whereas the third displays a median lifespan of ~30 weeks (Brown et al., 2015), highlighting the variability associated with differences in genetic background, gene targeting strategies, or C-terminal tags (see Table 1.1 for more detailed information). Using these mice, the partial loss-of-function phenotypes associated with the R306C mutation was linked to a disruption in protein interaction with the NCoR-SMRT co-repressor complex, resulting in MeCP2 failing to recruit NCoR-SMRT to chromatin and mediate transcriptional repression (Lyst et al., 2013). Thus these mutant mice provided critical molecular insights into RTT, highlighting loss of transcriptional repression by this complex as being critical to disease onset. The relatively mild phenotypic severity among R306C mice and human patients suggests that loss of transcriptional repression by NCoR-SMRT may be necessary, but not sufficient, to mediate the full severity of RTT phenotypes.

In this respect, nonsense mutations in the R270X and G273X mice are particularly interesting. Boys with the MeCP2 R270X mutation exhibit severe symptoms typical of *Mecp2* loss of function in human males, including neonatal encephalopathy and death before 4 years of age (Kankirawatana et al., 2006; Venâncio et al., 2007). However, boys with the MeCP2 G273X mutation live significantly longer with a relatively mild display of neuropsychiatric symptoms (Ravn

et al., 2003). The strong phenotypic differences associated with the difference in three amino acids is recapitulated in *MECP2* R270X and G273X mutant mice, which exhibit median lifespans of ~12 and 27 weeks, respectively (Baker et al., 2013). Further molecular investigations using these mice found that although both mutations abolished transcriptional repressor activity, the R270X mutation disrupted an AT-hook domain that severely impaired MeCP2's ability to recruit the nucleosome remodeler ATRX, cluster nucleosomes, and form higher order structures from nucleosomal arrays; in contrast, this AT-hook domain was maintained with the G273X mutation, which exhibited more moderate impairments for the same features (Baker et al., 2013). Thus, molecular differences in MeCP2's ability to modulate chromatin structure and mediate transcriptional repression likely underlie the differences in severity between these two mutations in both mice and humans.

1.4 UNDERSTANDING THE MOLECULAR FUNCTION OF MECP2

1.4.1 Synaptic Deficits Associated with MeCP2

Without question, the defining feature of the vertebrate brain is its ability to learn and adapt to changes in its external environment. This occurs through activity-dependent biochemical processes and structural dynamics in the neuron, resulting in widespread morphological changes at neuronal dendrites and synapses. Dendrites, which encompass many postsynaptic synapses, are highly specialized and extraordinarily complex cellular compartments that "branch out" to receive external excitatory or inhibitory sensory input and help inform an individual neuron as to whether it should propagate sensory stimuli to surrounding neurons. The ability to scale an individual neuron's electrical capacity and activity in response to dynamic changes in overall network activity encompasses the essence of synaptic plasticity, and has broad implications for the way that the brain processes and stores information in a higher-order manner.

Synaptic plasticity, which has primarily been studied in excitatory pyramidal neurons, is affected by both translational and transcriptional mechanisms (Guan et al., 2002; Holtmaat and Svoboda, 2009). Neuronal depolarization results in the intracellular release of calcium, whose

cytoplasmic concentration is relatively low under basal conditions. This results in the activation of multiple signaling pathways mediated by protein kinases (CaMKII, CaMKIV, PKA, MAPK, ribosomal S6 kinase) and protein phosphatases (calcineurin) that can modulate the activity of translational and transcriptional regulators (Ebert and Greenberg, 2013). Translational regulation enacts synaptic plasticity on a more immediate timescale through activation of mTOR and ERK signaling pathways, resulting in the local translation of previously inert mRNA transcripts that are localized to dendrites and encode various synaptic proteins; thus, *de novo* protein synthesis at the synapse will rapidly influence the compositional structure and function of postsynaptic structures and modulate signal input (Holtmaat and Svoboda, 2009). Neuronal activity also regulates transcriptional programs that affect the development and plasticity of the synapse. The most well described mechanism for activity-dependent transcriptional activation is observed with cyclic-AMP-dependent-responsive-element binding protein (CREB) activity. Upon phosphorylation by CaMKIV, CREB binds to specific responsive elements at gene promoters and recruits CREB-binding protein (p300/CBP), a histone acetyltransferase that post-transcriptionally modifies histones and alters chromatin structure to promote gene activation through the recruitment of RNA polymerase II (Ebert and Greenberg, 2013; Guan et al., 2009). Transcriptional regulation can thus be tailored in response to neuronal activity and regulate the abundance of mRNAs localized to the synapse, leading to dynamic long-term changes in synaptic plasticity.

MeCP2-deficient neurons in RTT patients and *Mecp2*-null animal models consistently demonstrate neuronal phenotypes that are characteristic of a loss of synaptic plasticity. Symptom onset among human RTT patients occurs during the first year of life when experience-dependent neuronal plasticity is guiding synapse development (Alvarez-Saavedra et al., 2007; Chahrour and Zoghbi, 2007; Chen et al., 2015). Morphologically, pyramidal neurons in post-mortem RTT patient brains, as well as IPSC-derived RTT patient neurons both show a striking reduction in soma size, dendritic arborization, disorganized afferent axonal fibers, and reduced dendritic spine density, consistent with the volumetric decrease of patient brains (Belichenko et al., 1994; Chapleau et al.,

2009; Kim et al., 2011; Kishi and Macklis, 2010; Marchetto et al., 2010; Murakami et al., 1992; Reiss et al., 1993). These phenotypes are recapitulated in pyramidal neurons of *Mecp2*-null mice, and can be induced through shRNA-mediated downregulation of MeCP2 expression in hippocampal slice cultures and through *in utero* neuronal transfections of wild-type mice (Belichenko et al., 2008, 2009; Chapleau et al., 2009; Landi et al., 2011; Wood et al., 2009). This suggests that neuronal deficits in plasticity are cell autonomous, which is supported by transplants of *Mecp2*-null mutant neurons into the brains of wild-type mice, which fail to rescue cellular deficits (Kishi and Macklis, 2010).

These morphological deficits are undoubtedly accompanied by abnormal electrophysiological responses in RTT neurons. Electrophysiological studies in *Mecp2*-null mice have shown that excitatory synapse strength is reduced concurrently with reductions in glutamatergic synapse number, and patch clamp recordings in the somatosensory cortex reveal decreased excitatory synaptic drive and increased inhibitory drive onto layer V pyramidal neurons (Chao et al., 2007; Dani et al., 2005). These experiments demonstrate that RTT cortical pyramidal neurons ultimately become hypoexcitable, as they fire less frequently and respond less robustly to excitatory input, which can disrupt overall network activity in the brain (Goffin et al., 2012, 2014; Nguyen et al., 2012). The shift in balance from excitation to inhibition in the brain has led the field to propose that MeCP2 functions to maintain experience-dependent maturation and regulation of neuronal circuits. This is supported by both the observation that MeCP2 is phosphorylated in an activity-dependent manner (Cohen et al., 2011; Ebert et al., 2013; Zhou et al., 2006), by the induction of RTT-like behavioral and cellular phenotypes upon conditional postnatal deletion of MeCP2 adult mice (Cheval et al., 2012; McGraw et al., 2011), and by the partial rescue of RTT-like phenotypes in mice upon conditional expression of MeCP2 in post-mitotic neurons of *Mecp2*-null animals (Alvarez-Saavedra et al., 2007; Giacometti et al., 2007; Guy et al., 2007; Kerr et al., 2012; Luikenhuis et al., 2004).

Although MeCP2 deficiency is associated with penetrant neuroanatomical features that are conserved across post-mortem human RTT patients, *Mecp2* mutant mice, and cultured

neurons, studies have also observed a wide degree of variability in the degree to which each neuron is affected (Belichenko et al., 2008, 2009; Wang et al., 2013). Much of this variability may be attributed to differences in genetic background, individual brain regions, or cell types, which suggests that individual neurons may be differentially affected by loss of MeCP2 function.

1.4.2 MeCP2 Interactions with Chromatin

The conserved structure of the methyl-CpG binding domain among MBD family members, coupled with a transcriptional repression domain and its strong enrichment at pericentromeric chromatin in the nucleus, supports the hypothesis that MeCP2 binds to methylated DNA and mediates transcriptional repression *in vivo* (Nan et al., 1996, 1997, 1998). As a result, there has been much interest in identifying target genes that are de-repressed in the absence of MeCP2 function and likely contribute to the etiology of RTT by determining where MeCP2 binds to chromatin *in vivo*. Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) has led to the functional identification of multiple DNA-binding proteins and the biological processes that they participate in (Furey, 2012). However, despite extensive study using ChIP-seq, the mechanism by which MeCP2 interacts with chromatin *in vivo*, as well as its precise localization throughout the neuronal genome, remains poorly understood. MeCP2 is highly abundant in the brain and is thought to coat the genome; the amount of MeCP2 molecules per neuron is sufficient to saturate all methyl-CpGs in the genome, and tracks DNA methylation genome-wide with a particular affinity for high densities of methylated CpGs (Baubec et al., 2013; Chen et al., 2015; Cohen et al., 2011; Skene et al., 2010). The high nuclear abundance of MeCP2 in post-mitotic neurons, together with the broad distribution of MeCP2 throughout the genome via ChIP-seq and its ability to affect higher-order nucleosomal structure, has led to the reassessment that MeCP2 may function as a global regulator of chromatin structure, rather than as a classical sequence-specific transcription factor that function at specific genes (Skene et al., 2010).

Biochemical assays have consistently demonstrated that MeCP2 preferentially binds to symmetrically methylated CpGs adjacent to A/T rich sequences *in vitro*, and prior studies traditionally assessed MeCP2 repressor function through binding to methylated promoters *in vivo* (Klose et al., 2005; Nan et al., 1997; Yu et al., 2000). RTT mutations that disrupt the MBD also confirm that the recognition of methylated CpGs by a functional MBD remains the primary determinant of MeCP2 binding *in vivo* (Baubec et al., 2013). However, recent observations from a number of studies question whether methyl-CpGs alone are sufficient to mediate MeCP2 interactions with chromatin in a physiological context. In addition to methylated CpGs, studies have found that methylated DNA in a non-CpG context (methyl-CpH) is highly enriched in the brain and comprises ~40-50% of methylated cytosines in the mature cortical neurons relative to other somatic cell types (Lister et al., 2013; Mo et al., 2015). Notably, DNMT3A deposits CpH methylation early in postnatal neuronal development, and the increase in methyl-CpH levels correlate with rates of synaptogenesis in the brains of humans and mice (Lister et al., 2013). Interestingly, MeCP2 expression also rapidly increases during the same developmental period (Chen et al., 2015; Kishi and Macklis, 2004; Shahbazian et al., 2002b). Methyl-CpH is found throughout the gene body and is associated with either transcriptional repression or low levels of expression, and patterns of methyl-CpH vary between individual neuronal cell types and correlate with cell type-specific gene expression in the mouse brain (Lister et al., 2013; Mo et al., 2015). In recent years, MeCP2 has been closely associated with the role of methyl-CpH in post-mitotic neuronal maturation through direct binding to methylated CpH, primarily in a methyl-CpA context, to mediate transcriptional repression (Chen et al., 2015; Gabel et al., 2015; Guo et al., 2014). Furthermore, methyl-CpA dinucleotides are enriched for higher MeCP2 ChIP-seq read density and are positively correlated with gene length, and as a result are throught to mediate the length-dependent repression of long genes observed across various *Mecp2*-mutant animal and cell culture models (Gabel et al., 2015).

In contrast, MeCP2 has also been shown to bind to hydroxymethylated CpGs in the mouse cerebellum, which is known to be associated with active expressed genes, particularly in a

cell type-specific context (Mellén et al., 2012). However, subsequent MeCP2 ChIP-seq experiments have been unable to reproduce or detect MeCP2 enrichment at hydroxymethylated cytosines *in vivo*, in contrast with methylated DNA in a CpG/CpH context (Gabel et al., 2015). MeCP2 may also associate with genomic regions in a methylation-independent manner. One study recently implicated MeCP2 in binding to GC-rich promoters of the genome in mouse olfactory epithelial tissue, regardless of the cytosine methylation status (Rube et al., 2016). This is in line with an early report that performed a genome-wide analysis of MeCP2 binding in the human neuronal SH-SY5Y cultured cell line and found that 63% of MeCP2-bound promoters were actively expressed genes, whereas only 6% of promoters were methylated (Yasui et al., 2007). Together, these studies argue against MeCP2 in silencing methylated promoters. Additional evidence reveals that, within murine embryonic stem cells where MeCP2 is ectopically overexpressed, the global loss of CpG methylation results in a redistribution of the protein towards regions with open accessible chromatin, including regions of increased DNAseI hypersensitivity, histone H3K4me1 and H3K27ac (Baubec et al., 2013). Although MeCP2 does exhibit methylation-independent binding to DNA that may be mediated by its C-terminal domain, it has been hypothesized that these interactions are non-specific and may occur *in vivo* as a means of searching for methylated regions of the genome (Hansen et al., 2010; Heckman et al., 2014). However, it remains an open question whether these marks are reproducible and representative of *Mecp2* binding *in vivo*, especially in the face of biochemical studies demonstrating MeCP2's strong preference for methylated DNA.

Although it is entirely plausible that MeCP2 function varies depending on the surrounding chromatin structure and methylation status, more research is required to determine where and how MeCP2 binds to chromatin *in vivo*, as evidence from current studies probing MeCP2 localization to chromatin currently remain in conflict.

1.4.3 Post-translational Modifications of MeCP2

Given the high abundance of MeCP2 in the brain, its broad histone-like genomic distribution, and the synaptic deficits associated with RTT, MeCP2 has been proposed to modulate chromatin structure and gene expression in an activity-dependent manner through the use of post-translational modifications (PTMs), similar to histone tail PTMs. Differential phosphorylation of MeCP2 has been shown to modulate MeCP2 protein interactions, affinity for methylated DNA and chromatin, gene expression, and neuronal morphology (Cohen et al., 2011; Ebert et al., 2013; Li et al., 2011; Tao et al., 2009; Zhong et al., 2012; Zhou et al., 2006). Given the importance of protein-kinase signaling and phosphorylation in synaptic plasticity, subsequent studies have demonstrated that MeCP2 is phosphorylated on multiple residues in response to neuronal activity from excitatory synapses (Chen et al., 2003; Zhou et al., 2006). Serine 421 and Threonine 308 are both phosphorylated in response to neuronal activity. Serine 421 is phosphorylated as a product of CaMKII kinase signaling and is correlated with MeCP2-mediated de-repression of *Bdnf*, which ultimately acts to decrease the number of excitatory synapses and increase the number of inhibitory synapses (Bloodgood et al., 2013; Zhou et al., 2006). This may occur through reduced affinity for methyl-CpGs as observed using southwestern blots (Chen et al., 2003), although ChIP-seq did not reveal extensive changes in MeCP2 binding upon Serine 421 phosphorylation (Cohen et al., 2011). In contrast, Threonine 308 phosphorylation is required to disrupt the interaction between MeCP2 and the NCoR-SMRT co-repressor complex (Ebert et al., 2013). Activity-dependent phosphorylation of these two residues may negatively regulate MeCP2 function by synergistically mediating the de-repression of gene targets, which may be important for the establishment of proper neuronal connectivity in the brain. In support of this notion, the S421A mutation in mice, which abolishes Serine 421 phosphorylation, is associated with cognitive impairments, abnormal behavioral responses to novel experiences, and increased neuronal dendritic complexity, similar to what is observed with mice that overexpress MeCP2 (Cohen et al., 2011; Collins et al., 2004). In contrast to Serine 421 and Threonine 308, Serine 80 is a residue that is phosphorylated in resting neurons of the mouse brain and becomes

dephosphorylated in the presence of neuronal activity (Tao et al., 2009). Mutation at this residue (S80A) leads to decreased association with chromatin in resting neurons, which suggests that this residue positively regulates MeCP2 function (Tao et al., 2009).

MeCP2 has also been shown to exhibit other PTMs. Small Ubiquitin-like Modifier (SUMO) is a small protein involved in various cellular processes, including transcriptional regulation, protein stability, response to stress, and apoptosis (Hay, 2005). MeCP2 is SUMOylated in response to NMDA, IGF-1, or CRF-mediated neuronal activity in the rat brain, and has been shown to release CREB from a repressor complex in order to enhance *Bdnf* mRNA expression (Tai et al., 2016). Importantly, the re-expression of SUMO-modified MeCP2 in *Mecp2*-null neurons can rescue several behavioral and electrophysiological deficits in conditional *Mecp2*-null mice, implicating SUMOylation as being a negative regulator of MeCP2 function, together with phosphorylation at Serine 421 and Threonine 308 (Tai et al., 2016). MeCP2 also undergoes poly(ADP-ribosyl)ation *in vivo*, and the loss of poly(ADP-ribosyl)ation on MeCP2 is associated with an increase in both MeCP2 binding to chromatin and its ability to cluster pericentromeric heterochromatin in MEF cells, suggesting that poly(ADP-ribosyl)ation negatively regulates MeCP2-mediated chromatin structural changes (Becker et al., 2016).

1.4.4 MeCP2 and Transcriptional Profiling

Since 2002, multiple MeCP2 gene expression studies have been performed using whole brain or individual brain regions of *MeCP2*-null (Ben-Shachar et al., 2009; Chahrour et al., 2008; Gabel et al., 2015; Jordan et al., 2007; Kriaucionis et al., 2006; Nuber et al., 2005; Tudor et al., 2002; Urdinguio et al., 2008) or *Mecp2*-mutant mice bearing RTT missense mutations (Baker et al., 2013; Gabel et al., 2015). These studies predominantly use male mice because 100% of the neurons in these mice express the null *Mecp2* allele, in contrast to female mice that undergo random XCI. These studies revealed that MeCP2-mediated gene expression changes are extraordinarily subtle, as earlier studies using cDNA arrays either failed to identify gene expression changes in *Mecp2*-mutant mice, or identified relatively few genes that were

misregulated in aged symptomatic mice but not early asymptomatic mice. The majority of these gene expression studies applied > 1.5 fold change cutoffs; by reducing the fold change cutoff to > 1.15, hundreds to thousands of genes were found to be differentially expressed depending on the brain region profiled (Ben-Shachar et al., 2009; Chahrour et al., 2008). Although many differentially expressed genes (DEGs) were identified, it still remains unclear how small changes in gene expression among many individual genes lead to the morphological and physiological effects seen in *Mecp2*-null animals, although one hypothesis suggests that neurons may be sensitive to subtle changes in the dosage of mRNAs (Tudor et al., 2002). As a result, many gene expression studies have been limited to gene-specific interpretations and candidate gene experimental approaches. Although a small number of misregulated genes, such as the corticotropin-releasing hormone (*Crh*), opioid receptor mu-1 (*Oprm1*), and *Brain-derived neurotrophic factor* (*Bdnf*) are sufficient to mediate behavioral phenotypes when misregulated in mice (Samaco et al., 2012; Zhou et al., 2006), these genes tend to encode receptors or secreted hormones whose expression is known to dramatically influence organism physiology and behavior.

Furthermore, there is an extraordinary amount of variability among individual DEGs across multiple studies, mouse lines, and cultured cells. As a result, there are few validated target genes and no consistent molecular pathways that have been identified. One of the earliest cDNA microarray studies profiled gene expression changes in human *MECP2* wild-type and mutant fibroblasts and lymphoblastoid cells but found that MeCP2 did not cause reproducible changes in mRNA transcript levels, even across multiple clones that were derived from the same RTT patient (Traynor et al., 2002). Another study compared gene expression changes between two germline *Mecp2*-null mouse lines (Chen et al., 2001; Guy et al., 2001) using the cerebellum, which is mostly comprised of excitatory granule neurons and largely limits cellular heterogeneity. Although three post-natal time points were assessed, with the number of DEGs increasing with age, only 2.6% and 2% of all DEGs found in either mutant mouse line overlapped with the other mutant mouse line, despite the similarities in phenotype (Jordan et al., 2007). This study observed that

genetic background, as well as the gene targeting strategy, helps contribute to the large variation in the types of DEGs that are detected from various *Mecp2* mutant mice.

A portion of this variation may be due to the astounding complexity and cellular heterogeneity of the mammalian CNS. The brain in particularly comprises numerous anatomically distinct brain regions; each region possesses a myriad of intermixed cell types that differ in morphology, function, and electrophysiological properties. To accommodate this cellular diversity, different cell types establish distinctive gene expression profiles that are maintained by a complex interplay of cell type-specific transcription factors, epigenetic modifications (e.g. methylcytosine, posttranslational modification of histone tails), and alterations to chromatin accessibility and structure. Experiments that solely rely on individual brain regions may thus reflect composite profiles that easily obscure these cell type-specific differences and impede an accurate assessment of gene regulation and function. Indeed, recent studies that account for CNS heterogeneity have been able to identify gene expression markers for novel, functionally distinct neuronal subtypes and have clarified regulatory relationships between intragenic CG/CH methylation and gene expression (Fuzik et al., 2016; Mo et al., 2015; Tasic et al., 2016; Usoskin et al., 2015). These studies highlight the need for novel innovations that parse cellular heterogeneity in exchange for increased biological insight, especially in light of a neurodevelopmental disease context. In support of this view, it was recently found that manually sorted *Mecp2* mutant neurons displayed gene expression changes that were unique to each cell type and would have otherwise been undetected using whole tissue homogenates, which suggests that MeCP2 may modulate gene expression in a cell type-specific manner (Sugino et al., 2014). This contrasts with previous studies that found similar gene expression within the hypothalamus and cerebellum and concluded that MeCP2 regulates the same complement of genes across different brain regions (Ben-Shachar et al., 2009; Chahrour et al., 2008).

Alternatively, the variability in the identity of individual DEGs across different mutant mice or cell lines may also be a feature of MeCP2 loss-of-function and the RTT phenotype. Because

this variability, coupled with the high abundance and broad distribution of MeCP2 throughout the neuronal genome (Cohen et al., 2011; Skene et al., 2010), has made the identification of specific MeCP2 gene targets exceedingly difficult, researchers have proposed the alternative hypothesis that MeCP2 may play a global role in transcriptional regulation. Consistent with this hypothesis is a recent study which found that when genes are sorted by gene length and fold changes are averaged every 200 genes, MeCP2-deficient neurons display a length-dependent increase in gene expression specifically within the brain, implicating MeCP2 in mediating the repression of long genes enriched for methyl-CpA that are important for neuronal function (Gabel et al., 2015; Sugino et al., 2014). The most notable feature about this study is the conservation of this phenomenon across *Mecp2*-null or *Mecp2*-mutant primary neurons, brain tissue, and individual cell types from mice, human *MECP2* ESC-derived neurons, and RTT patient mutant fibroblasts, and that such effects can also be recapitulated in conditionally-deleted *Dnmt3a*-null neurons that are responsible for methyl-CpA deposition. This study provided strong evidence for MeCP2 in binding to long genes and mediating global transcriptional repression of long neuronal genes related to synaptic functions. However, a major question that remains is how the ectopic upregulation of long neuronal genes genome-wide leads to the reductions in glutamergic synapse number, reduced dendritic branch complexity, and cell autonomous hypoexcitability of neurons in RTT, aside from compromising general cellular health. In this respect, the length-dependent de-repression of long neuronal genes in the absence of MeCP2 is not completely sufficient to explain the occurrence of RTT phenotypes.

Finally, gene expression studies from human post-mortem RTT brain tissue reveal generally larger fold changes in gene expression, in contrast to gene expression changes in *Mecp2*-null male mice (Colantuoni et al., 2001; Tudor et al., 2002). Although such differences may be attributed to evolutionary differences between humans and mice, the majority of transcriptional and molecular studies rely on male hemizygous *Mecp2*-mutant mice. Given that RTT patients are predominantly female, it remains an open question whether gene expression changes in female heterozygous *Mecp2*-mutant mice, the most clinically relevant model for RTT,

will closely resemble those that have been observed in male hemizygous mutant mice. Transcriptional profiling in female heterozygous *Mecp2*-mutant mice may thus offer new insights into RTT pathogenesis and MeCP2 function.

1.4.5 Proposed Roles for MeCP2 Function

MeCP2 as a Transcriptional Repressor

Although *in vivo* evidence supports a role for MeCP2 in transcriptional repression by binding to methylated CpA and mediating length-dependent repression of long neuronal genes, these are relatively recent observations. Historical evidence supporting MeCP2 as a transcriptional repressor is largely derived from biochemical assays examining its ability to bind to methylated DNA and interact with transcriptional co-repressors both *in vivo* and *in vitro*. Transient transfection of MeCP2 in heterologous cells indicated that this protein could bind to methylated promoters and mediate long-range repression of reporter constructs, consistent with the role of a transcriptional repressor (Nan et al., 1997). It was later found that transcriptional repression by MeCP2 *in vivo* required the recruitment of the histone deacetylase co-repressor complex Sin3A via the MeCP2 TRD (Nan et al., 1998). This was followed by the identification of NCoR-SMRT co-repressor recruitment, whose stoichiometric interaction with MeCP2 can be regulated by phosphorylation on R306 in an activity-dependent fashion (Ebert et al., 2013; Lyst et al., 2013). Consistent with MeCP2 being a transcriptional repressor via HDAC co-repressor recruitment *in vivo*, the loss of *Mecp2* function is also associated with a global increase in histone H3 acetylation in post-mitotic neurons (Shahbazian et al., 2002a; Skene et al., 2010). Although the genomic locations of this mark and its correlation with gene expression changes is not yet understood, the increase in global histone acetylation was accompanied by an increase in spurious transcription of repetitive elements throughout the genome, which may implicate MeCP2 in suppressing transcriptional noise (Skene et al., 2010). However, given that transcriptional repression by MeCP2 is partially relived by trichostatin A, an inhibitor of histone deacetylase activity, it is

possible that MeCP2 mediates transcriptional repression independent of co-repressor recruitment (Nan et al., 1998; Yu et al., 2000).

MeCP2 has been shown to associate with nucleosomal arrays by binding to linker DNA between individual nucleosomes and mediating chromatin compaction, even in the absence of methylated DNA, consistent with a proposed role for MeCP2 in mediating transcriptional repression (Georgel et al., 2003; Nan et al., 1997). MeCP2-bound nucleosomal arrays exhibit reduced enzymatic digestion of micrococcal nuclease (MNase), indicative of reduced accessibility to increased chromatin compaction. Furthermore, exogenous MeCP2 expression and accumulation at DAPI+ nuclear chromocenters of Pmi28 mouse myoblasts is associated with a decrease in the number of chromocenters that are visualized in the nucleus, which suggests that MeCP2 can mediate the clustering of pericentromeric chromatin (Agarwal et al., 2011). The expression of MeCP2 bearing RTT-associated missense mutations was correlated with a decrease in MeCP2 accumulation and an increase in the number of chromocenters within the nucleus, suggesting that up to two-thirds of RTT-associated missense mutations that were assessed can impair the ability of MeCP2 to cluster pericentromeric chromatin *in vivo*. However, these observations are limited by the use of artificial nucleosomal arrays *in vitro*, which are performed under conditions that may promote non-specific binding of MeCP2 to DNA (Thambirajah et al., 2012). Likewise, the exogenous overexpression and detection of MeCP2 accumulation at pericentromeric chromatin only represents one type of methyl-CpG-rich chromatin structure *in vivo* that can be imaged by confocal microscopy. It remains unclear whether MeCP2 directly compacts neuronal chromatin *in vivo* in order to mediate transcriptional repression, particularly under physiological conditions and in the presence of other chromatin modifiers and epigenetic marks.

MeCP2 as a Transcriptional Activator

Several gene expression studies have supported a role for MeCP2 in mediating transcriptional activation. Previous studies identified MeCP2-mediated transcriptional changes in

the hypothalamus and cerebellum of *Mecp2*-null and *Mecp2*-overexpression mice under the assumption that true MeCP2 gene targets would show inverse changes between the two conditions (Ben-Shachar et al., 2009; Chahrour et al., 2008; Chen et al., 2015). They found that, although most gene expression changes were subtle, thousands of genes in the hypothalamus were misregulated by the loss or gain of MeCP2 protein expression, and 85% of misregulated genes were identified as being activated by MeCP2. Mechanistically, MeCP2 may mediate the transcriptional activation of these genes through synergistic interactions with the transcriptional activator CREB1, as both proteins were found to directly interact with each other and were found at the promoters of activated genes, but not repressed genes (Chahrour et al., 2008). Thus MeCP2 may promote the activation of genes through interactions with transcriptional activators.

By using RNA-spike ins to control for equivalent cell numbers among human ESC-differentiated post-mitotic neurons, another study reported that global reductions in transcription among *MECP2*-deficient neurons that worsened with time (2-4 weeks post-differentiation) (Li et al., 2013a). Genes that were downregulated were expressed at substantially higher levels than genes that were upregulated and exhibited higher ratios of hydroxymethyl-CpG/methyl-CpG, suggestive of deficits in active transcription within *MECP2*-mutant neurons. These transcriptional deficits were associated with morphological and physiological deficits observed in RTT patients and mouse models, including decreases in soma size, a reduction in dendritic branch complexity, and reduced action potential rates. Notably, these neurons also displayed reduced incorporation of radiolabeled methionine over a 30-minute period, consistent with a global reduction in protein translation rates. Many of the ribosomal proteins were also transcriptionally downregulated, consistent with an additional study that also identified reductions in nascent transcription of ribosomal RNAs using *Mecp2*-null ESC-derived neurons (Yazdani et al., 2012). Together, these studies implicate MeCP2 in mediating transcriptional activation for a large number of genes that are associated with the morphological and electrophysiological neuronal deficits observed in post-mortem human RTT patients and *Mecp2*-mutant mice. However, a general mechanism by which

MeCP2 may promote transcriptional activation to mediate proper neuronal morphology and function still remains unclear.

Studies performed *in vivo* using *Mecp2*-mutant mice have proven to be equally illuminating by revealing that MeCP2 modulates chromatin structure on a global scale, and these changes may be consistent with transcriptional activation on a genome-wide scale. A recent high-resolution imaging study in heterozygous *Mecp2*-null female mice reported that MeCP2-deficient nuclei in CA1 pyramidal neurons are associated with an increase in chromatin compaction (Linhoff et al., 2015). Interestingly, compaction in the absence of MeCP2 was only observed among pyramidal neurons that typically express high levels of MeCP2. Cerebellar granule cells, which have relatively smaller cell sizes and nuclear volumes, contain more heterochromatin, and exhibit lower levels of MeCP2 expression, did not display a similar increase in chromatin compaction. Chromatin compaction in MeCP2-deficient neurons was also associated with a global increase and redistribution in H4K20me3, a marker of transcriptional repression, and reduced S5 phosphorylation of RNA Polymerase II, a marker of transcriptional activity. This study suggests that cell type-specific abundances of MeCP2 may modulate chromatin structure in different cell types to promote transcriptional activity, lending support to MeCP2 functioning as a transcriptional activator.

MeCP2 also competes with linker histone H1 through both *in vitro* nucleosomal arrays and *in vivo* cellular FRAP assays, which may have global effects on the accessibility of chromatin *in vivo*. Linker histone H1 can promote higher-order aggregation of chromatin fibers that may be repressive (Bednar et al., 1998; Ghosh et al., 2008; Nan et al., 1997), and is traditionally considered a transcriptional repressor (Bustin et al., 2005; Fan et al., 2005; Nishiyama et al., 2012). Loss of MeCP2 chromatin binding in neurons leads to a doubling of linker histone H1 expression in neurons and may promote chromatin condensation (Ghosh et al., 2010; Ishibashi et al., 2008; Skene et al., 2010). The high expression of MeCP2 in the brain, coupled with the 50% decrease in linker histone H1 expression relative to other somatic tissues and cell types, may

promote more open and uncondensed chromatin at transcriptionally active genes. Additional evidence also supports MeCP2 in modulating neuronal chromatin structure that would be consistent with transcriptional activation. Biochemical fractionation and isolation of rat brain nuclear lysates reveals that most MeCP2 is enriched in soluble fractions of chromatin and is found on nucleosome-bound DNA, suggesting that MeCP2 binds to more euchromatic and nuclease-accessible regions of the genome (Thambirajah et al., 2012). Similarly, MNase digestion of isolated cerebellar nuclei from *Mecp2*-null mice revealed a subtle decrease in the accessibility of 5hmC-enriched chromatin when compared to wild-type mice, indicating that MeCP2 promotes more nuclease-accessible chromatin structure at transcriptionally active genes (Mellén et al., 2012).

Additionally, although MeCP2 exists as a monomeric protein in solution, MeCP2-bound nucleosomal arrays form higher order structures mediated by MeCP2-MeCP2, MeCP2-DNA-MeCP2, or DNA-MeCP2-DNA intermolecular contacts, which is consistent with studies that report that MeCP2 mediates three-dimensional chromatin loops *in vivo* (Bednar et al., 1998; Klose and Bird, 2004; Nikitina et al., 2007a, 2007b). Notably, MeCP2-dependent compaction and higher-order organization of nucleosomal arrays are also facilitated by domains outside of the MBD, suggesting that other domains of MeCP2, including AT-hooks and the C-terminal tail, may also interact with chromatin independent of a methylcytosine context (Baker et al., 2013; Heckman et al., 2014; Nikitina et al., 2007b). In the neonatal brain, MeCP2 recruits the nucleosome remodeler ATRX to imprinted gene loci such as *H19/Igf2* and *Gtl2/Dlk1*, and remodels nucleosomes near these loci to promote a more open, de-condensed chromatin structure (Kernohan et al., 2010, 2014). This allows linker DNA regions to be extended in length, exposing CTCF binding motifs that help to maintain three-dimensional chromatin conformations required for proper gene regulation at these imprinted loci. Although evidence supporting MeCP2 mediating three dimensional chromatin conformations in adult mice has been limited at best (Horike et al., 2005), the observation that ATRX is expressed in adult mice and is mislocalized to chromatin in *Mecp2*-mutant mice suggests that three-dimensional interactions may occur *in vivo*.

Other Proposed Mechanisms for MeCP2

MeCP2 has been shown to bind mRNA and siRNA with high affinity comparable to that of methylated DNA, although these binding events are mutually exclusive (Jeffery and Nakielny, 2004). MeCP2 has thus been implicated in mediating alternative splicing (Maunakea et al., 2013; Young et al., 2005), as an alternative hypothesis to explain the subtle gene expression changes in the absence of MeCP2 function. Alternative splicing is critical for brain function, as various transmembrane ion channels utilize various cell type-specific isoforms that are responsive to different neuronal voltages and conductance in the plasma membrane. MeCP2 has been shown to affect the activity of splicing mini-genes in HeLa cells, and changes in alternative splicing were identified for a subset of genes in the cerebral cortex of *Mecp2*$^{308/y}$ mice (Young et al., 2005). MeCP2 also directly interacts with Y box-binding protein 1 (YB1), a conserved RNA splicing factor that forms a major component of messenger ribonucleoprotein particles (Young et al., 2005). Another study that performed MeCP2 ChIP-seq in IMR90 cells found MeCP2 to be enriched at alternatively spliced exons in a methyl-CpG dependent fashion, particularly those exons that are included in the mature transcript (Maunakea et al., 2013). MeCP2 knockdown using siRNAs, as well as trichostatin A treatment, results in the loss of MeCP2 binding at alternatively spliced exons, leading to increased histone acetylation and an increase in the frequency of alternative exon skipping. This supports a role for MeCP2 in binding to alternatively skipped exons and maintaining local histone hypoacetylation to promote alternative exon inclusion into the final mRNA transcript.

	Mutation	Protein Product	Molecular Effect	Symptom onset	Median Lifespan	Gross Phenotypes	Cellular/Molecular Phenotypes	Reference
MeCP2 Null Mice	Exon 3 Deletion (mixed background)	Lacks full MBD C-terminal protein fragment detected	No targeting to methylated DNA	Males: 4-6 weeks Females: ~4 months	Males: 10 weeks Females: N/A		Increased cell packing, reduced soma size, decreased dendritic arbor complexity	Chen et al. 2001
	Exon 3 & 4 Deletion (C57BL/6 background)	No protein detected	Null allele	Males: 4-6 weeks Females: ~3 months	Males: 8 weeks Females: N/A		Increased cell packing, reduced soma size, decreased dendritic arbor complexity	Guy et al. 2001, Pelka et al. 2006
MeCP2 RTT Mutation (Targeted *Mecp2* gene locus)	**R106W-Tavi** (C57BL/6 background)	Impaired MBD, TRD preserved	Impaired binding mCpG Reduced protein (~20%)	Males: 4-6 weeks Females: ~3 months	Males: 10 weeks Females: N/A			This study
	R133C-GFP (C57BL/6 background)	Impaired MBD, TRD preserved	Impaired binding mCpG Reduced protein (55%)	Males: 6 weeks Females:~6 months	Males: 42 weeks Females: N/A	Gradual onset of phenotypes		Brown et al. 2015
	T158A (C57BL/6 background)	Impaired MBD, TRD preserved	Impaired binding mCpG Reduced protein (~60%)	Males: 5 weeks Females: 17 weeks	Males: 16 weeks Females: > 6 months	Reduced brain weight, body weight, motor & learning and memory deficits (not as severe as null), anxiety	Reduced soma size	Goffin et al. 2012
	T158M-GFP (C57BL/6 background)	Impaired MBD, TRD preserved	Impaired binding mCpG Reduced protein (~30%)	Males: 6 weeks Females: 3-4 months	Males: 13 weeks Females: N/A	Rapid onset of phenotypes		Brown et al. 2015
	T158M-Tavi (C57BL/6 background)	Impaired MBD, TRD preserved	Impaired binding mCpG Reduced protein (~30%)	Males: 4-6 weeks Females: ~3 months	Males: 14 weeks Females: N/A			This study
	R168X (C57Bl/6 background)	C-terminal truncation Retains MBD, lacks TRD	Truncation prior to NLS No nuclear localization	Males: 4 weeks Females: ~9 months	Males: 12 weeks Females: N/A	Similar to *Mecp2*-null mice		Lawson-Yuen et al. 2007, Schaevitz et al., 2013
	R255X (C57Bl/6 background)	C-terminal truncation Retains MBD, lacks TRD	Truncation prior to NLS, No protein product detected	Males: 6 weeks Females: 17 weeks	Males: 9 weeks Females:	Similar to *Mecp2*-null mice	LTP impairments mTorc1 pathway abnormalities	Pitcher et al. 2015
	R306C-GFP (C57BL/6 background)	Preserved MBD, TRD impaired	Disrupts NCoR interaction, No effect on protein stability (compared to WT-GFP control)	Males: 6 weeks Females: 3-4 months	Males: 30 weeks Females: N/A	Initial rapid onset, phenotypically stable through 33 weeks, becomes more severe		Brown et al. 2015
	R306C (No tag) (C57Bl/6 background)	Preserved MBD, Impaired TRD	Disrupts NCoR interaction	Males: 6 weeks Females: N/A	Males: 18.5 weeks Females: N/A			Lyst et al. 2015
MeCP2 RTT Mutation (Human PAC/BAC, Transgenic alleles expressed in *Mecp2*-null background)	R111G-EGFP (Human PAC) (FVB background)	Impaired MBD TRD preserved	Impaired binding mCpG, No effect on stability (in mice with endogenous and transgenic MeCP2 expression)	Males: 4-6 weeks Females: N/A	Males: 11 week Females: N/A			Heckman et al. 2014
	R270X (Human PAC) (FVB background)	C-terminal truncation, disrupts AT-hook	Impaired repression, Severe impairement to facilitate oligomerization of nucleosomal arrays	Males: ~6 weeks Females: N/A	Males: 12 weeks Females: N/A	More rapid symptom onset, appears similar to *Mecp2*-null mice	Severe ATRX mislocalization by 7wk	Baker et al. 2013
	G273X (Human PAC) (FVB background)	C-terminal truncation, preserves AT-hook	Impaired repression, Intermediate impairement to facilitate oligomerization of nucleosomal arrays	Males: Females: N/A	Males: 27 weeks Females: N/A	More mild symptoms and gradual onset	Partial ATRX mislocalization by 9wk	Baker et al. 2013
	R306C-EGFP (Human PAC) (FVB background)	Preserved MBD, TRD Impaired	Disrupts NCoR interaction, No effect on stability (in mice with endogenous and transgenic MeCP2 expression	Males: 6 weeks Females: N/A	Males: 18 weeks Females: N/A		Partial ATRX mislocalization by 7wk	Heckman et al. 2014
MeCP2 Non-RTT Mutations	Targeted floxed *Mecp2* (Used to generate Jaenisch Exon 3 deletion above)	Hypomorphic expression of WT protein (~50% reduction)		Males: Females:	Males: Females:			Chen et al. 2001, Phenotype described in Kerr et al. 2008, Samaco et al. 2005
	A140V (Mouse BAC) (C57BL/6NCrL background)	Preserved MBD & TRD	Disrupts ATRX interaction	Males: Females:	Males: Females:		Increased cell packing, decreased dendritic arbor complexity	Jentarra et al. 20010
	MeCP2 C-term Truncation (C57BL/6J, 129/SvEv, mixed)	Truncation after aa 308 Preserved MBD & TRD		Males: Females:	Males: Females:	Tremor, myoclonic seizures, kyphosis, stereotyped forelimb movements, enhanced anxiety		Shahbazian et al. 2002
	S80A	Preserved MBD & TRD	Impaired activity-dependent phosphorylation	Males: N/A Females: N/A	Males: N/A Females: N/A			Tao et al. 2009
	S421A	Preserved MBD & TRD	Impaired activity-dependent phosphorylation	Males: N/A Females: N/A	Males: N/A Females: N/A	Abnormal social and novel object responses	Increased dendritic arbor complexity, E/I balance shifted towards inhibiton	Cohen et al. 2011
	S421A;S424A	Preserved MBD & TRD	Impaired activity-dependent phosphorylation	Males: N/A Females: N/A	Males: N/A Females: N/A	Better cognitive performance compared to WT mice	Increased LTP and synaptogenesis	Tao et al. 2009
MeCP2 Transgenic Mice	Human *MECP2* BAC Transgene	WT protein, ~1-fold expression		Males: 10 weeks Females:	Males: 1 year Females:	Hypoactivity and seizures after 20 weeks	Increased LTP and synaptogenesis	Collins et al. 2004

Table 1.1: List of *Mecp2* mouse models made to date

	CRE-Line	Affected Neuronal Population	Additional Phenotypes	Premature Lethality?	Hypoactivity	Altered Weight	Kyphosis	Repetitive Behavior	Self-injury	Hindlimb Clasping	Motor incoordination	Ataxia	Anxiety	Social interaction	Aggression	Acoustic startle	Prepulse inhibition	Premature lethality	Breathing dysfunction	Learning & Memory	Seizures	Study
Constitutive Null	N/A	Central Nervous System		~10 weeks of age	■	■	■			■	■	■	■					■	■			Chen et al. 2001, Guy et al. 2001
Floxed *Mecp2* (hypomorph)	N/A	Central Nervous System		~10 weeks of age		■					■	■	■	■		■			■			Samaca et al. 2008, Chao et al. 2010
Inhibitory Neurons	Viaat-Cre	All GABAergic & glycinergic interneurons			■	■	■	■	■	■	■	■		■	■	■	■	■	■	■		Chao et al. 2010
	Dlx5/6-Cre	Forbrain GABAergic interneurons	Network deficits in auditory ERP			■		■			■	■		■		■	■				■	Chao et al. 2010, Goffin et al. 2014
	PV-Cre	Parvalbumin+ GABAergic interneurons	Abolishes critical period plasticity in V1 cortex	29 weeks of age							■	■		■		■				■	■	He et al. 2014, Ito-Ishida et al. 2015
	SOM-Cre	Somatostatin+ GABAergic interneurons		35 weeks of age				■													■	Ito-Ishida et al. 2015
Excitatory Neurons	Camkllα-Cre	Forebrain Glutamatergic neurons & Glia	Limited behavioral abnormalities			■				■	■	■	■	■								Chen et al. 2001, Gemelli et al. 2006
Neuromodulatory Neurons	TH-Cre	Dopaminergic & Noradrenergic neurons	Motor abnormalities		■						■											Samaco et al. 2009
	Pet1-Cre	Serotonergic neurons													■							Samaco et al. 2009
Hypothalamic Neurons		POMC+ neurons in Hypothalamus	Over-eating and obesity																			Wang et al. 2014
	Sim1-Cre	SIM1+ neurons in Hypothalamus	Heightened stress, hyperphagia and obesity			■							■		■							Fyffe et. al 2008
	Viralyl-mediated	Basolateral Amygdala																		■		Adachi et al. 2009
	HoxB1-Cre	Brainstem and spinal cord	Abnormal respiration, heart rate, lethality	13-18 weeks of age							■											Ward et al. 2011

Table 1.2: List of *Mecp2* mouse conditional studies and associated phenotypes

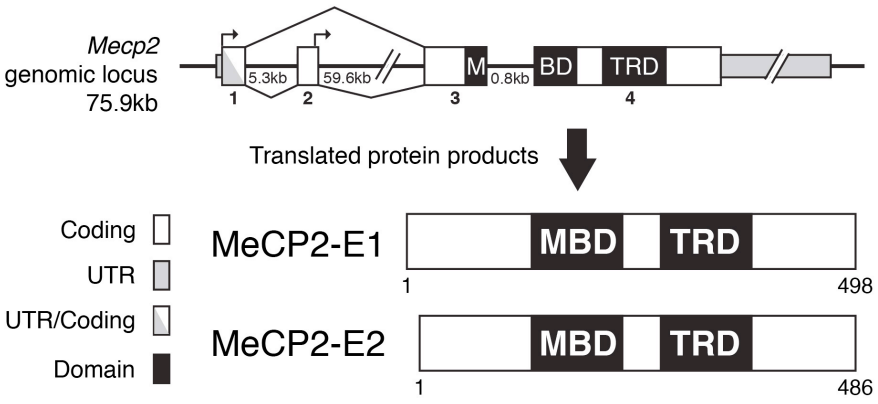

FIGURE 1.1: Murine *Mecp2* gene locus, alternative isoforms, and protein products

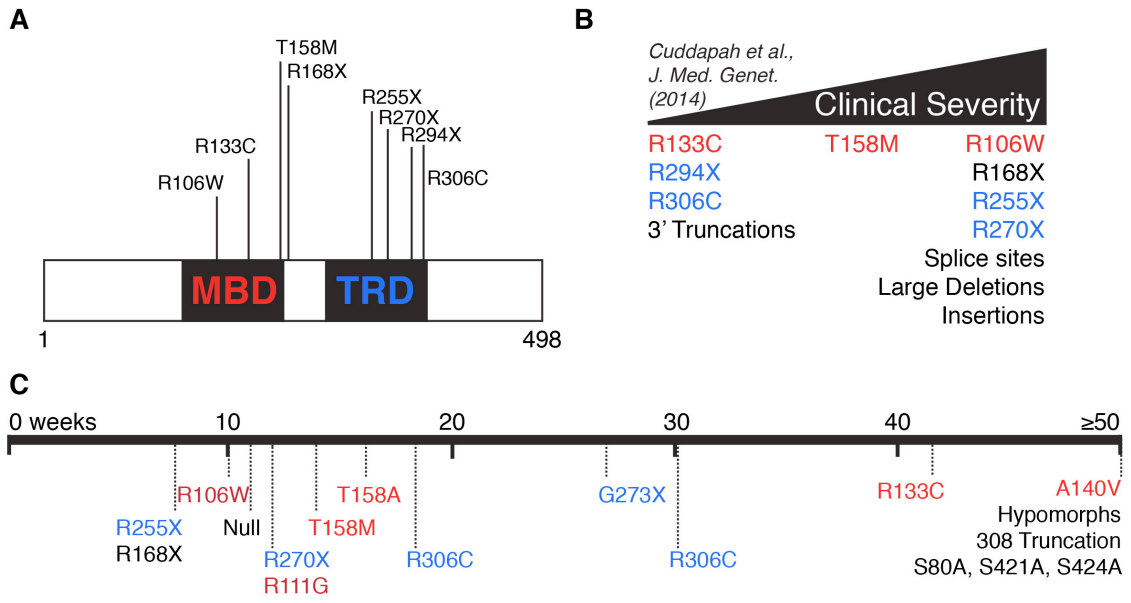

FIGURE 1.2: Clinical and phenotypic severity of RTT-associated point mutations

(A) Locations of the eight most frequent MeCP2 missense and non-sense mutations associated with RTT. MBD, Methyl-CpG Binding Domain; TRD, Transcriptional Repression Domain; NEO, Neomycin cassette.

(B) Human RTT patient clinical severity associated with RTT-missense mutations depicted in **(A)**.

(C) Phenotypic severity among mouse models bearing RTT-asssociated missense mutations. Depicted is the median survival of various mouse models in weeks.

CHAPTER 2: Loss of MeCP2 is associated with distinct expression changes in the striatum, hypothalamus, and cerebellum

Adapted from: Zhao, Y.*, Goffin, D.*, Johnson, B.S.*, Zhou, Z. (2013) Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. *Neurobiol. Dis*. 59: 257-266.

ABSTRACT

Rett syndrome (RTT) is a neurodevelopmental disorder characterized by developmental regression beginning 6–18 months after birth, followed by a lifetime of intellectual disability, stereotyped behaviors, and motor deficits. RTT is caused by mutations in the gene encoding MeCP2, a methyl-CpG binding protein believed to modulate gene transcription. Gene expression studies of individual brain regions have reported that Mecp2 loss-of-function leads to both activation or repression of its gene targets in mice. Conditional deletion of MeCP2 from different brain regions has revealed unique insights into the role of these structures in mediating particular RTT-like phenotypes. However, the function of MeCP2 in the striatum, a major brain region involved in motor control and executive cognitive functions, has yet to be studied. Here, we characterized the gene expression changes in the striatum of Mecp2 mutant mice. We found a number of differentially expressed genes in the striatum of both constitutive Mecp2-null mice and mice lacking MeCP2 only from forebrain GABAergic neurons. These changes only occurred when MeCP2 expression levels had reached mature levels and RTT-like symptoms were manifest, supporting a role for MeCP2 in maintaining proper brain function. Many of the gene expression changes identified in the striatum have not previously been shown to change in the hypothalamus or cerebellum. Bioinformatic analysis of differentially expressed genes in striatum as well as hypothalamus and cerebellum revealed that loss of MeCP2 does not affect the global landscape of gene expression. Additionally, we uncovered a number of differentially expressed genes in the liver of Mecp2-null mice suggesting an important role for MeCP2 in non-neuronal tissues. Collectively, our data suggest that the differential expression of genes following loss of MeCP2

occurs in a tissue, or cell-type specific manner and thus MeCP2 function should be understood in a cellular context.

INTRODUCTION

RTT is a severe X-linked neurological disorder that affects 1 in 10,000 – 15,000 females. Patients with RTT experience a relatively normal first 6-18 months of life after which they enter a period of developmental stagnation or regression. Subsequently, patients lose learned motor and communication skills, and develop hand stereotypies and gait apraxia (Katz et al., 2012; Neul et al., 2010). Approximately 95% of RTT cases are associated with mutations in the gene encoding methyl-CpG binding protein 2 (MeCP2).

Genetic studies using Mecp2 knockout mice, or mice carrying RTT-associated mutations, have established a causal role of MeCP2 in the etiology of RTT (Chen et al., 2001; Goffin et al., 2012; Guy et al., 2001; Pelka et al., 2006; Shahbazian et al., 2002a). Conditional deletion of Mecp2 from the mouse brain recapitulates the phenotypes observed in Mecp2-null mice such as reduced brain and body weight, decreased locomotor activity, motor incoordination, gait and breathing abnormalities (Chen et al., 2001; Guy et al., 2001). These studies demonstrate that RTT is primarily a neurological disorder and that MeCP2 is indispensable for proper brain function. Furthermore, conditional deletion of Mecp2 from specific brain regions or specific population of brain cells have demonstrated that different symptoms of RTT may manifest through loss of MeCP2 function from particular brain regions or cell types (Adachi et al., 2009; Chao et al., 2010; Fyffe et al., 2008; Lioy et al., 2011; Samaco et al., 2009; Ward et al., 2011). However, while specific neurotransmitter systems and neuronal populations play important roles in the etiology of individual RTT-like phenotypes, the large-scale disruption of MeCP2 function is required for the appearance of phenotypes that mimic constitutive Mecp2-null mice (Goffin and Zhou, 2012).

MeCP2 is a highly abundant nuclear protein that binds methylated CpGs (5mC) in vitro and in vivo (Skene et al., 2010). Through its interaction with co-repressor complexes, such as Sin3A and histone deacetylases 2 (HDAC2), MeCP2 was originally postulated to repress gene

transcription (Nan et al., 1997); however initial gene expression studies did not find significant changes in gene expression upon loss of MeCP2 (Kriaucionis et al., 2006; Tudor et al., 2002). To reduce the potentially confounding effects of cellular heterogeneity in the brain, other studies examined gene expression changes using individual brain regions; these studies identified hundreds to thousands of genes whose expression is affected by MeCP2 dysfunction in the hypothalamus and cerebellum (Ben-Shachar et al., 2009; Chahrour et al., 2008). Notably, most of these differentially expressed genes were down-regulated in Mecp2-null mice and up-regulated in Mecp2-overexpressing mice, arguing against a primary role for MeCP2 in gene repression. This additional role for MeCP2 in gene activation may result from the recruitment of CREB to MeCP2 (Chahrour et al., 2008) or the binding of MeCP2 to 5-hydroxymethylcytosine (5hmC), which is enriched over active genes (Mellén et al., 2012; Szulwach et al., 2011). Since the pattern of 5mC and 5hmC localization across the genome is cell-type dependent (Mellén et al., 2012), those genes identified as differentially expressed in the hypothalamus and cerebellum may reflect the composite changes across multiple cell types.

One brain region that has received little attention in the RTT field is the striatum. This brain region serves as the input center for the basal ganglia, a subcortical region of the brain that processes and integrates information associated with motor control and various executive cognitive functions. Indeed, hand stereotypies, impaired motor coordination and disrupted cognitive functions are common for all RTT patients. In addition, RTT patients often display a volumetric decrease in the size of the striatum (Reiss et al., 1993; Subramaniam et al., 1997). Deletion of MeCP2 from forebrain GABAergic neurons leads to the expression of several RTT-like phenotypes (Chao et al., 2010). However, the role the striatum plays in mediating these phenotypes is not known. Additionally, the genes regulated by MeCP2 in the striatum have yet to be defined. Given that 97-98% of striatal neurons are GABAergic medium spiny neurons (MSNs), this cell homogeneity affords this region a unique advantage for the study of MeCP2-regulated gene expression.

In this study, we profiled gene expression changes in the striatum in mice following the conditional deletion of MeCP2 from the striatum during early development and at an adult stage. We identified 21 differentially expressed genes only in adult mice and not during development, highlighting the importance of MeCP2 in maintenance of proper brain function. We also examined gene expression changes in the striatum of constitutive Mecp2-null mice and found that 59 genes are significantly down-regulated and 68 genes are significantly up-regulated in the absence of MeCP2. We found that the majority of differentially expressed genes in the striatum are distinct from those identified in the hypothalamus and cerebellum. Weighted gene co-expression network analysis revealed that the overall transcriptome differs largely between brain regions rather than between WT and Mecp2-null genotypes, indicating that the gene expression landscape is not globally affected by loss of MeCP2. Together, our data suggests that MeCP2 regulates gene expression in a region-specific manner, most likely driven by cell type, and that the aggregated differences in gene expression within individual cells and regions lead to the pathogenesis of RTT. Given that MeCP2 is a 5mC and 5hmC binding protein and these methylation patterns likely differ between different cell types, we propose that the role of MeCP2 in regulating target gene expression should be studied with individual cell types in consideration.

MATERIALS AND METHODS

Mouse strains, Sample collection, and RNA extraction

Experiments were conducted in accordance with the ethical guidelines of the National Institutes of Health and with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania. Conditional deletion of Mecp2 from forebrain GABAergic neurons was achieved by breeding $Mecp2^{flox/+}$ females (Chen et al., 2001) to transgenic Dlx5/6-cre males. Constitutive Mecp2-null mice (Guy et al., 2001) were generated by breeding $Mecp2^{-/+}$ females to WT males. All mice were maintained on the C57BL/6 background. Striata were dissected from cKO mice at P7 and P90 and immediately frozen in liquid nitrogen. Striatal and liver tissue were dissection from constitutive Mecp2-null mice at P60 and flash frozen in liquid nitrogen. RNA was extracted using TRIzol reagent (Invitrogen Corporation, Carlsbad, CA, USA), DNase I treated,

and purified using the miRNAeasy mini kit according to the manufacturer's protocol (Qiagen, Valencia, CA, USA).

Microarray Target Preparation and Hybridization

Microarray services were provided by the Penn Microarray Facility, including quality control tests of the total RNA samples by Agilent Bioanalyzer and Nanodrop spectrophotometry. All protocols were conducted as described in the Affymetrix GeneChip Expression Analysis Technical Manual. Briefly, 100 ng of total RNA was converted to first-strand cDNA using reverse transcriptase primed by poly(T) and random oligomers that incorporated the T7 promoter sequence. Second-strand cDNA synthesis was followed by in vitro transcription with T7 RNA polymerase for linear amplification of each transcript, and the resulting cRNA was converted to cDNA, fragmented, assessed by Bioanalyzer, and biotinylated by terminal transferase end labeling. cRNA yields and 5 µg of labeled cDNA was added to Affymetrix hybridization cocktails, heated at 99ºC for 5 min and hybridized to Affymetrix Mouse Exon 1.0 ST GeneChips (for constitutive Mecp2-null mice and their WT littermates) or Affymetrix GeneChip Mouse Gene 1.0 ST Arrays (for cKO mice and their WT littermates) according to manufacturers instructions.

Microarray Data Analysis

The bioinformatics analysis of all the microarray datasets used in this manuscript was processed in the MAC terminal window and R environment. The raw CEL files of hypothalamus (accession number GSE11150) and cerebellum (accession number GSE15574) datasets were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database. Four methods were used to process the raw exon array data and to calculate gene-level expression values. The Affymetrix Power Tools (APT) command line program was used to process raw data using the robust multichip average method and generate gene-level expression values based on the core probesets using median polish. The JETTA bioconductor package was used to normalize raw data using the quantile method and to calculate gene-level

expression values based on the core probesets using median polish. The last two methods utilized the affy and xmapcore (exonmap) bioconductor packages. Raw data was processed using the robust multichip average method. Probesets that mapped uniquely to gene exons using the exmapcore package were used for subsequent analysis. Gene-level expression values were calculated using either the median or mean value of all uniquely-mapped exon probsets. Differentially expressed genes were identified using the limma bioconductor package. A linear model was fitted for every gene, and the p-values of differentially expressed genes were calculated using the empirical Bayes moderated t-test. Next, multiple comparisons were processed to adjust the raw p-value of t-test using Benjamini and Hochberg's method and to generate the false discovery rate (FDR). The significance threshold was set at a fold change ≥1.2 with a FDR ≤ 0.05 for identifying differentially expressed genes between Mecp2-null mice and their WT littermates. The significance threshold was set at a fold change ≥2 with a FDR ≤ 0.05 for tissue and region-enriched genes between striatum, hypothalamus, cerebellum, and liver datasets. Genes that were identified by at least one of the four methods described were considered to be differentially expressed genes.

The GOEAST online software using default settings was used for Gene Ontology (GO) enrichment analysis. Gene ID conversion was performed using the DAVID Bioinformatics Resources.

Weighted Gene Co-expression Network Analysis (WGCNA)

The weighted gene co-expression network was built using the WGCNA package in the R environment. The JETTA package was used to generate the gene expression profiles of WT and constitutive Mecp2-null mice, and a network was constructed using all striatum, cerebellum, and hypothalamus datasets. Network constructions were performed using the blockwise Modules method in the WGCNA package. Using the scale-free topology criterion, the power of six was chosen to construct the total network. Dynamic hybrid tree cutting was used to cut the hierarchical clustering trees and to identify individual modules. The minimum module size was set

to 30 genes and the minimum height for merging modules was set to 0.1. Default parameters were used for other settings.

Quantitative Real-time PCR

For each sample, 1 µg of total RNA was digested with DNAseI (Invitrogen, Carlsbad, CA) and reverse-transcribed by oligodT-priming using SuperScriptIII reverse transcriptase (Invitrogen, Carlsbad, CA). The amount of each transcript of interest present in the sample was measured by quantitative real-time PCR on 1/10th of the resulting cDNA using SYBR Green detection (Applied Biosystems) on an ABI 7900HT instrument. Each sample was measured in triplicate using exon-spanning qRT-PCR primers. GAPDH amplification served as an internal RNA control for relative quantification.The following mouse primers were used for qRT-PCR analysis: Gapdh (F: 5'-AATGGTGAAGGTCGGTGTG-3'; R: 5'-GTGGAGTCATACTGGAACATG-3'), Mecp2 (F: 5'-CATACATAGGTCCCCGGTCA-3'; R: 5'- CAGGCAAAGCAGAAACATCA-3'), Irak1 (F: 5'-GTCTTGGATAGCCTGCAACTG-3'; R: 5'-TGAGGGATTTGTCAGAGTGAAC-3'), Drd3 (F: 5'-TTATCCACATCCCTGAAGCTG-3'; R: 5'-ACAGTGGGTATTAAGAACGTGAG-3'), Exph5 (F: 5'-TCGGTCATCCTTTGCTTCAC-3'; R: 5'-AAGGTCTTACACGCTCATCG-3'), Satb1 (F: 5'-TCTAGGAAGAGGAAGGCTTGG-3'; R: 5'-CGCAGAAAACTGGTAACATGG-3'), Dsg1c (F: 5'-GGAGCTTAGAGTTAGAGTGATGG-3'; R: 5'-CAAGTTATTTGGCTCATCGGC-3'), Dlk1 (F: 5'-TGTCAATGGAGTCTGCAAGG-3'; R: 5'-ATTCGTACTGGCCTTTCTCC-3'), Robo3 (F: 5'-TTGTGACTAAGCCCCAGAAC-3'; R: 5'-CTGACTAGGGAAAAGCAGGAC-3').

Data Deposition

The datasets that were generated and reported in this paper have been deposited into the National Center for Biotechnology Information GEO database through accession number GSE42895.

RESULTS

Differentially expressed genes in the striatum of Mecp2 conditional knockout mice

The striatum serves as the major site of input and integration for cortical, thalamic, and midbrain afferents in the basal ganglia, a brain region essential for maintaining proper sensorimotor and executive cognitive functions. The manifestation of motor and cognitive disturbances in RTT, together with the observed decreased striatal volume in RTT patients (Reiss et al., 1993; Subramaniam et al., 1997), point to a possible role for the striatum in the etiology of RTT. A causal role for the striatum in the pathogenesis of RTT, or a biochemical maker of striatal dysfunction, however, has yet to be examined.

Given the potential role for MeCP2 in the repression and activation of gene transcription, we examined whether loss of MeCP2 alters gene expression within the striatum. The striatum affords a unique advantage for gene expression studies because it is composed primarily of GABAergic MSNs. This homogeneity in cell type reduces the confounding effects that multiple unique cell types with highly divergent gene transcriptional programs can have on gene expression analyses (Heiman et al., 2008). Since altered neuronal activity caused by the loss of MeCP2 from afferent inputs to the striatum may indirectly affect striatal gene expression, we conditionally deleted MeCP2 from forebrain GABAergic neurons, including striatal MSNs, while preserving expression in those neurons that provide inputs to the striatum. We achieved this by breeding floxed Mecp2 mice ($Mecp2^{2lox/+}$) (Chen et al., 2001) to mice expressing Cre-recombinase localized to striatal MSNs and forebrain GABAergic interneurons (Dlx5/6-cre) (Monory et al., 2006). The behavioral consequences of deletion of MeCP2 from forebrain GABAergic neurons using Dlx5/6-cre have previously been determined (Chao et al., 2010). These mice do not show any of the overt RTT-like phenotypes – such as hypoactivity, gait, breathing difficulties or premature lethality – that are readily apparent in constitutive Mecp2-null mice. However, behavioral characterization of these mice revealed phenotypic alterations involving social interaction preference, motor coordination and repetitive behaviors (Chao et al., 2010). To determine the efficiency of cre recombinase in the striatum of these mice, we performed immunohistochemistry using an antibody directed against MeCP2 in the male progeny of mice with the $Mecp2^{lox/y}$; Dlx5/6-cre genotype (referred hereafter as cKO) and their $Mecp2^{2lox/y}$

(WT) littermates. Indeed, we demonstrate a significant loss of MeCP2 immunoreactivity from the striatum of cKO mice but not their WT littermates (Figure 2.1A). In addition, close examination of MeCP2 expression revealed the existence of <1% of cells in the striatum (Figure 2.2).

RTT-like symptoms in mice manifest in an age-dependent manner. Mice show relatively normal behavior during the early stages of postnatal development followed by the gradual appearance and increased severity of RTT-like phenotypes. We therefore examined striatal gene expression levels at two developmental time points: the first at P7 during early postnatal development when MeCP2 protein levels are still increasing and the second following development at P90 when MeCP2 protein levels have plateaued (Skene et al., 2010). In addition, behavioral testing has previously shown that cKO mice exhibit motor deficits at this time point (Chao et al., 2010). We therefore dissected the striatum from male cKO and WT littermates at P7 and P90 and obtained genome-wide whole transcript coverage expression data using Affymetrix GeneChip Mouse Gene 1.0 ST Arrays (Gene Arrays). We first examined whether striatal gene expression changed over the course of development from P7 to P90. Using the criteria of a fold-change of ≥ 1.2 with FDR ≤ 0.05 we identified over 7,000 genes that were differentially expressed between P7 and P90 in both WT and cKO samples (Figures 2.1B-C). These data demonstrate that there is a large-scale reorganization of the striatal transcriptome across development and that this occurs to a similar extent in both WT mice and mice lacking MeCP2 from striatal neurons.

We also compared gene expression changes between cKO mice and their WT littermates at P7 and P90. During early postnatal development at P7 we did not detect any significant differential gene expression between genotypes using the aforementioned criteria (Figure 2.1D). In contrast, in adult mice at P90 we identified the differential expression of 21 genes in cKO mice relative to WT littermates (Figure 2.1E, Table 2.1). These results demonstrate that the conditional loss of MeCP2 from striatal neurons leads to changes in the expression of a small number of genes only following postnatal development. Therefore, MeCP2 may not be required for striatal gene regulation during the early stages of development, but rather plays a critical role in maintaining proper brain function. In support of this hypothesis, conditional deletion of Mecp2

from adult mice is sufficient to recapitulate a numerous RTT-like symptoms (Cheval et al., 2012; McGraw et al., 2011; Nguyen et al., 2012) and restoration of MeCP2 in adult mice can reverse RTT-like neurological phenotypes (Guy et al., 2007).

Differentially expressed genes in the striatum of Mecp2 constitutive knockout mice

The limited number of differentially expressed genes in the striatum of cKO mice is in contrast to previous studies examining gene expression changes in the hypothalamus or cerebellum of constitutive Mecp2-null mice where hundreds to thousands of genes are differentially expressed (Ben-Shachar et al., 2009; Chahrour et al., 2008). This difference may result from differences in cellular composition between brain regions, the identity of the mouse model used, the array platform used or the cell autonomous and non-cell autonomous effects of MeCP2. To address the possibility that the identity of the brain region mediates these differences, we examined gene expression changes in the striatum of the same constitutive Mecp2-null mice using the same Affymetrix GeneChip Mouse Exon 1.0 ST Arrays (Exon Array) used in previous studies (Ben-Shachar et al., 2009; Chahrour et al., 2008). Striata were dissected from symptomatic P60 Mecp2-null mice (Guy et al., 2001) and their WT littermates, and total RNA was extracted, converted to cDNA and hybridized to individual exon arrays. Since exon arrays have 1.2 million probe sets with approximately 40 probes per gene, the calculation of gene expression levels is a greater challenge than that required for gene arrays: that is, multiple probes will be mapped to individual genes with varying intensities, leading to difficulty calculating the level of gene expression. Multiple distinct, but related, methods have been developed to analyze exon array data with respect to the calculation of gene expression levels. To analyze differentially expressed genes in a comprehensive manner we applied four different methods to measure gene expression levels: assessing expression level using the mean values from individual probes for a given gene, the median values from individual probes for a given gene, the Affymetrix Power Tools (APT) software package and the JETTA software package (Seok et al., 2012).

Using these different methods, we have compiled four lists of genes that are differentially expressed in the striatum of WT and Mecp2-null mice. These four lists of genes overlap with one other to a large degree, but unique genes are also identified by each method. Using criteria of fold change of ≥ 1.2 with FDR ≤ 0.05 we identified a total of 127 genes that are differentially expressed in the striatum of Mecp2-null mice in at least one of the four approaches (68 up-regulated genes, 59 down-regulated genes) (Supplementary Dataset S1). Of these, 67 were identified using at least two of these analytical methods (Supplementary Dataset S1) and 56% of those genes dysregulated in P90 cKO mice were also differentially expressed in Mecp2-null mice (Table 2.1). Importantly, we find that Mecp2 is significantly down-regulated, whereas interleukin-1 receptor-associated kinase 1 (Irak1) is significantly up-regulated in Mecp2-null mice. Since the Irak1 gene is located only 3 kb distal to Mecp2, this alteration is likely due to the altered genomic environment caused by deletion of Mecp2 exons 3 and 4 (Urdinguio et al., 2008).

To verify the exon array analysis, we randomly selected six genes from our list of differentially expressed genes (Exph5, Robo3, Drd3, Satb1, Dsg1c, and Dlk1) and analyzed their mRNA levels in WT and Mecp2-null striatal tissues using quantitative real-time PCR (qRT-PCR). We also included Mecp2 and Irak1 as positive controls. We found that all eight of these selected genes show expression changes that are consistent with those seen in our exon array data (Figure 2.3A), validating our approach in using exon array analysis to identify differentially expressed genes in the striatum of Mecp2-null mice.

To characterize the major biological themes present in differentially expressed striatal genes, we assigned Gene Ontology (GO) annotations to the genes and tested for significant enrichment of functionally related terms (Figures 2.3B-C). As expected, most of the enriched terms involve neural morphology or function. We found that the 59 down-regulated genes are significantly enriched with functions associated with ion and chemical homeostasis (Figure 2.3B), whereas the 68 genes up-regulated in Mecp2-null mice are highly enriched for functions associated with transcriptional regulation, development and differentiation (Figure 2.3C).

Therefore, the differentially expressed genes identified from the striatum are likely functionally related to neuronal function.

Differentially expressed genes in the striatum are distinct from other brain regions

To determine whether gene expression changes upon loss of MeCP2 are consistent across the brain regions, we compared the identified 127 differentially expressed striatal genes with genes previously shown to be significantly altered by loss of MeCP2 (defined as those genes showing fold change of ≥ 1.2 with FDR ≤ 0.05) in the hypothalamus (Chahrour et al., 2008) or cerebellum (Ben-Shachar et al., 2009) between WT and Mecp2-null mice. Unsupervised hierarchical clustering of differentially expressed genes between Mecp2-null and WT littermates in these three brain regions revealed largely distinct classes of genes affected by the loss of MeCP2 (Figure 2.4A). Comparison of those differentially expressed genes revealed that only a small number of genes that are altered in all three brain regions (Figure 2.4 and Supplementary Dataset S2). Indeed, only 2 genes (including Irak1) are commonly up-regulated between brain regions (Figure 2.4B), whereas 3 genes (including Mecp2) are commonly down-regulated (Figure 2.4C). In addition, comparison between differentially expressed genes in the cerebellum of Mecp2-null mice and their WT littermates identified using RNA-seq (Mellén et al., 2012) revealed only three genes were commonly down-regulated (including Mecp2) in the striatum. Furthermore, analysis of the GO terms for these differentially expressed genes yields a transcriptional signature for the striatum that is different to that observed in the hypothalamus or cerebellum (Figure 2.3B-C) (Ben-Shachar et al., 2009; Chahrour et al., 2008). These data support therefore that MeCP2 regulates gene expression in a region-specific manner.

Similar gene co-expression networks in WT and Mecp2-null mice

Given the findings that MeCP2 binds to 5mC and 5hmC across the genome (Mellén et al., 2012; Skene et al., 2010), it has been suggested that MeCP2 may affect global gene transcription in addition to specific genes. To assess the global effect of MeCP2 on gene

transcription, we analyzed gene expression level data from striatum, hypothalamus and cerebellum in Mecp2-null mice and their WT littermates using hierarchical clustering algorithms. We found that the overall gene expression profiles are clustered based on brain regions but not genotypes (Figure 2.5A). These data demonstrate that the overall biochemical signature of an individual brain region is maintained in the absence of MeCP2.

To gain higher-order and system level insight into how loss of MeCP2 affects the expression of functionally related genes, we next applied weighted-gene co-expression network analysis (WGCNA) to gene expression data from striatum, cerebellum and hypothalamus. WGCNA examines groups of genes whose expression profiles are highly correlated across samples and gives rise to gene modules containing subsets of genes with similar biochemical and functional properties as well as anatomical localization (Johnson et al., 2009; Oldham et al., 2008; Zhang and Horvath, 2005). This technique has recently been used to identify modules of dysregulated genes in autism patients, providing evidence for convergent molecular pathologies in ASD (Voineagu et al., 2011). We therefore constructed a co-expression network using the entire data set, composed of both Mecp2-null mice and their WT littermate samples from striatum, hypothalamus and cerebellum. Gene expression profiles were classified into nine modules (ME1 - 9) with the expression levels of each module summarized by the first principle component (the module eigengene).

We first asked whether global differences in the organization of the brain transcriptome exist between different brain regions. Indeed, eigengene values across all modules were markedly different between the striatum, hypothalamus and cerebellum (Figures 2.5B and 2.6), consistent with the existence of robust modules of co-expressed genes that are related to specific cell types and reflect anatomical localization and functional divergence (Oldham et al., 2008). We then sought to identify discrete groups of co-expressed genes showing transcriptional differences between Mecp2-null mice and their WT littermates. We found no significant differences in module eigengenes between WT and Mecp2-null mice across the striatum, hypothalamus or cerebellum (Figures 2.5B and 2.6). Therefore, the loss of MeCP2 function does not alter the general

organization and molecular distinctions of transcriptomes in different brain regions. Finally, we wanted to determine whether those genes differentially expressed in the striatum were distributed across modules or clustered together. We found that 70% of differentially expressed genes in the striatum are clustered within three modules: 33% of genes within module 1, 25% in module 3 and 12% in module 4 (Figure 2.5C). These data suggest therefore that loss of MeCP2 from striatum leads to alterations in genes that are functionally related to one another. The overall transcriptome differs largely between brain regions rather than between WT and Mecp2-null genotypes.

Evaluation of gene expression variation in Mecp2-null mice

MeCP2 has been suggested to organize global chromatin structure and tighten transcriptional regulation by dampening transcriptional noise (Muotri et al., 2010; Skene et al., 2010). This raises the possibility that loss of MeCP2 function may lead to the stochastic dysregulation of genes within a given brain region across individual animals: that is, altered gene expression profiles within different brain regions of Mecp2-null mice may reflect an increase in inter-subject variance in gene expression. To examine this possibility we calculated the variance for individual gene expression levels across animals in the striatum, hypothalamus and cerebellum of Mecp2-null mice and their WT littermates. However, we found no significant differences in gene expression variance in any of these three brain regions (Bartlett's test, Figure 2.7). To investigate the abundance of gene expression may override individual variation across the transcriptome, we further classified all expressed genes into three groups based on their expression levels (low, medium or high), and calculated expression variance within these individual groups. Again, we found no significant differences in gene expression variance between genotypes in any of these groups (Figure 2.8). These data indicate that increased variance in gene expression levels is not the major contributor to the differential expression of genes between Mecp2-null mice and their WT littermates. Therefore, loss of MeCP2 function is unlikely to modulate gene expression levels through a stochastic disruption of gene regulation

processes; rather MeCP2 appears to modulate gene expression through a coordinated, cell-type specific mechanism. Consistent with this, we found that MeCP2 also affects neuronal morphology in a brain region- and cell type-specific manner (Wang et al., 2013).

Loss of MeCP2 function affects both neuronal and non-neuronal genes

Distinct gene expression changes occurring in different brain regions would be expected if those genes were particularly enriched in a single brain region. To investigate this possibility, we isolated liver tissue from Mecp2-null mice and normalized gene expression data from striatum, hypothalamus and cerebellum to that of liver in order to identify region-enriched genes. We next generated a classification scheme by a direct comparison of WT gene expression profiles between our liver dataset and the datasets from striatum, hypothalamus, and cerebellum (Figure 2.9A). Genes exhibiting higher fold expression (fold change ≥ 2 with FDR ≤ 0.05) in the three brain regions relative to liver were considered neuronal-enriched genes. Among these, those that only appeared in a single brain region relative to the remaining two were considered specific brain region-enriched genes. Using these criteria, we identified striatum- (622 genes), hypothalamus- (1,109 genes) and cerebellum-enriched genes (714 genes) (Figure 2.9A and Supplementary Dataset S3). To verify the fidelity of this classification scheme, we obtained the top 50 highly expressed genes in the striatum from the Allen Mouse Brain Atlas and confirmed that 98% of the genes from the atlas are present in our list of striatum-enriched genes (Supplementary Dataset S4). A similar analysis with the previously published hypothalamus- and cerebellum-enriched gene lists revealed 83% and 89% overlap, respectively.

We next examined whether genes differentially expressed between WT and Mecp2-null mice were classified as brain region-enriched genes or demonstrated similar expression levels in different brain regions. We found that the majority of genes deregulated in Mecp2-null datasets tended to be those expressed across different brain regions or showed similar expression levels in the liver. Indeed, only a small portion, 3 to 10%, of all differentially expressed genes, are enriched in a single brain region (Figure 2.9A). Thus, brain region-specific differential gene

expression in Mecp2-null mice is unlikely to be a consequence of the enrichment of those genes in that particular brain region. Rather, other tissue-specific factors likely control this separation in differential gene expression.

MeCP2 is ubiquitously expressed throughout the body although its function in non-neuronal tissues is not well understood (Shahbazian et al., 2002b). To examine whether loss of Mecp2 function affects gene expression in non-neuronal tissues, we profiled gene expression levels from the liver of Mecp2-null mice and their WT littermates. We found that 161 genes are differentially expressed in the livers of Mecp2-null compared to their WT littermates (Figure 2.9B and Supplementary Dataset S5). Of these genes, 73 are up-regulated and were enriched for lipid, fatty acid and small molecule metabolism as determined by GO analysis (Figure 2.9C). Since MeCP2 expression is relatively low in the liver (Shahbazian et al., 2002b), these expression changes may be indirectly associated with MeCP2 due to general alterations in metabolism not yet characterized in Mecp2-null mice. Nevertheless, these data suggest that Mecp2-null mice may have altered liver metabolism consistent with a recent finding that mice lacking MeCP2 show numerous alterations in liver metabolism (Pitcher et al., 2013). It is possible therefore that these altered metabolic processes may also contribute to the pathology of RTT.

DISCUSSION

RTT is a devastating neurological disorder that leads to the manifestation of a number of diverse symptoms including gait apraxia, loss of purposeful hand movements, respiratory abnormalities and loss of communication. The particular symptoms observed in RTT suggest that striatal dysfunction may play an important role in its etiology. Indeed, volumetric analyses of RTT individuals have consistently identified a decrease in the size of the neostriatum, even when taking into account the smaller brain size of these patients (Reiss et al., 1993; Subramaniam et al., 1997). The striatum serves as the input center for the basal ganglia, a subcortical region of the brain which processes and integrates information associated with learning and memory, motor control and various executive cognitive functions. The conditional deletion of MeCP2 from forebrain GABAergic neurons, including MSNs of the striatum, leads to the development of a

number of symptoms in mice including impaired motor coordination, ataxia and social deficits (Chao et al., 2010). The role that MeCP2 loss-of-function in the striatum plays in the development of these symptoms, however, is unclear. Similarly, the identity of genes affected by loss of MeCP2 in this region has not been investigated.

In order to examine changes in gene expression within the striatum we utilized two mouse models: mice lacking MeCP2 only from forebrain GABAergic neurons including MSNs of the striatum, and constitutive Mecp2-null mice. Using the first model, we profiled gene expression changes in the striatum across development from P7 to P90. We found considerable changes in the gene expression profiles of the striatum during these time points, with over 7,000 genes being affected developmentally in both cKO mice and their WT littermates. In contrast, we observed only a small number of differentially expressed genes between cKO mice and their WT littermates. These alterations only occurred at P90 when MeCP2 levels are high and not during early postnatal development. These data suggest therefore that MeCP2 is not required for striatal gene regulation during early development; rather, MeCP2 appears to be essential for maintaining proper brain function at later postnatal stages. Indeed, conditional deletion of Mecp2 from adult mice is sufficient to recapitulate many RTT-like symptoms and altered hypothalamic gene expression (Cheval et al., 2012; McGraw et al., 2011; Nguyen et al., 2012).

Notably, the number of differentially expressed genes in striatum is markedly different in cKO striatum compared to those previously observed in hypothalamus and cerebellum of constitutive Mecp2-null mice (Ben-Shachar et al., 2009; Chahrour et al., 2008). To address the possibility that these differences were the result of different mouse models we repeated our striatal gene expression studies using symptomatic Mecp2-null mice and exon arrays similar to previous studies. Indeed, up to 127 genes were differentially expressed in symptomatic Mecp2-null mice at P60 compared to WT littermates. The greater number of genes identified in the striatum of Mecp2-null mice relative to cKO mice may result from differences in cell autonomous and non-cell autonomous effects of MeCP2 loss-of-function. Most of the genes identified in cKO

mice were also identified in Mecp2-null mice suggesting that they are indeed regulated by MeCP2.

Among the small number of differentially expressed genes found in the striatum, the majority of the affected genes were not identified in analyses of the hypothalamus or cerebellum. These data suggest that MeCP2 regulates gene expression in a region- or cell-type-specific manner. Furthermore, the majority of these region-specific changes in gene expression involve genes that are typically expressed in multiple brain regions and non-neuronal cell types. Distinct gene expression profiles in different brain regions are likely governed by local alterations to chromatin structure in addition to cell-type specific 5mC and 5hmC localization (Mellén et al., 2012; Szulwach et al., 2011). Thus region-specific alterations in gene expression may reflect differential binding of MeCP2 to 5mC and 5hmC, respectively.

Our analysis of the gene expression changes in the striatum revealed considerably fewer genes differentially expressed by loss of MeCP2 relative to the hypothalamus and cerebellum that together exhibited > 2,000 genes differentially expressed (Ben-Shachar et al., 2009; Chahrour et al., 2008). The larger number of genes altered in these two regions may reflect the greater heterogeneity in cell types within these regions in comparison to the striatum that is primarily composed of MSNs. It would be interesting to determine whether individual cell types, such as Granule cells, Purkinje cells and Bergmann glia from the cerebellum, demonstrate a more restricted alteration in gene transcription in Mecp2-null mice. Moreover, analysis of gene expression at a systems level using WGCNA revealed modules of genes functionally and anatomically classified based on brain region and are unaltered by loss of MeCP2. However, those genes that are differentially expressed in the striatum cluster within discrete sets of modules arguing that these genes are functionally related to one another. These findings suggest that MeCP2 does not alter the gene expression landscape globally, but rather shows a brain region- or cell-type-specific effect. Thus, while individual gene expression changes in individual cell types may be small, the composite effects, however, lead to global disruption in brain function and the pathogenesis of RTT.

Most of the work regarding MeCP2 has centered around its function in the brain since brain-specific deletion of MeCP2 is sufficient to recapitulate most RTT-like symptoms observed in constitutive Mecp2-null mice (Chen et al., 2001; Guy et al., 2001). Although MeCP2 is ubiquitously expressed, its function in other tissues remains unknown. In this study, we also examined gene expression changes in the liver of symptomatic Mecp2-null mice and identified 161 genes that were differentially expressed, with 73 upregulated genes being enriched for lipid, fatty acid and small molecule metabolism. These data may indicate that Mecp2-null mice may have altered liver metabolic processes. Indeed, it has been suggested that Mecp2-null mice have a metabolic disorder (Pitcher et al., 2013). Mecp2-null were found to have marked increases in sera levels of triglycerides and cholesterol and decreased sera levels of insulin-like growth factor binding-protein 2 (Pitcher et al., 2013). These alterations were associated with an increase in fat, but not lean mass and insulin resistance (Pitcher et al., 2013). Since MeCP2 expression remains low in the liver (Shahbazian et al., 2002b), it is not known whether these alterations in the expression of genes related to metabolism in the liver are directly caused by loss of MeCP2 function in the liver, or are secondary to occur indirectly through altered metabolic pathways in Mecp2-null mice. Further study into the function of MeCP2 in the liver and the greater characterization of metabolic processes affected by loss of MeCP2 will provide greater insight into the role of metabolism in the pathogenesis of RTT.

In summary, we identified a number of differentially expressed genes in the striatum of Mecp2-null mice. These differentially expressed genes in the striatum were largely unique to those previously identified in the cerebellum or hypothalamus. We have therefore uncovered a previously under-appreciated feature of MeCP2 in regulating gene expression in a region-specific manner within the brain. We demonstrate that differential gene expression in different regions of the Mecp2-null brain do not occur due to region-specific expression or increased gene expression variance or within a particular functional biochemical network. Differential gene expression is likely caused by other region-specific mechanisms such as local differences in 5mC and 5hmC

patterns. Our findings add new possibilities to the molecular function of MeCP2 and identify genes that may be novel therapeutic targets for the treatment of RTT.

CHAPTER 2 FIGURES

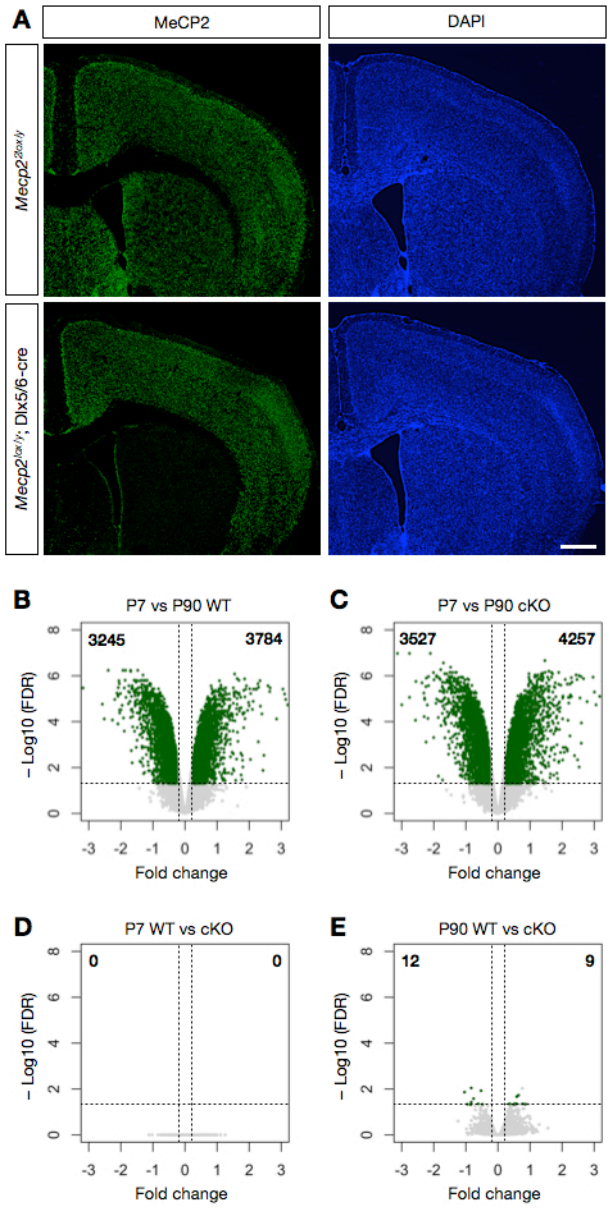

FIGURE 2.1: Gene expression changes in the striatum of *Mecp2* conditional knockout mice. (A) Immunohistochemistry for MeCP2 expression in *Mecp2*$^{2lox/y}$ and *Mecp2*$^{lox/y}$; Dlx5/6-cre mice at P90. The presence of Dlx5/6 leads to a substantial loss of MeCP2 from striatum. Scale bar corresponds to 500 µm.

(B) Volcano plots depicting fold-change and FDR-correct p-values for striatal gene expression between P7 versus P90 in WT and *Mecp2*$^{2lox/y}$; Dlx5/6-cre (cKO) mice.

(C) Volcano plots for gene expression changes between WT and cKO mice at P7 and P90. Green dots represent genes showing a fold change of ≥ 1.2 with FDR ≤ 0.05 that were considered differentially regulated

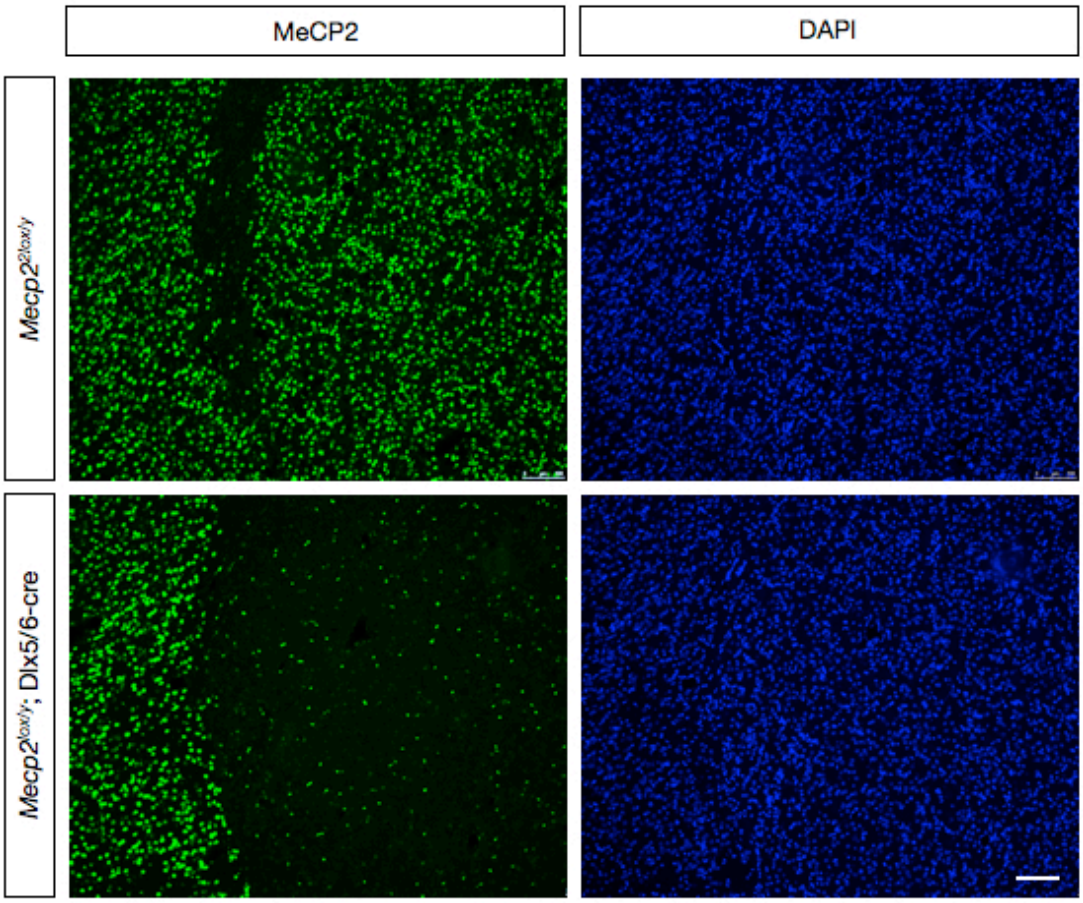

FIGURE 2.2: Loss of MeCP2 from the striatum with expression of Dlx5/6-cre

Immunohistochemistry for MeCP2 expression in *Mecp2*$^{2lox/y}$ and *Mecp2*$^{lox/y}$*; Dlx5/6-cre* mice at P90. The presence of Dlx5/6 leads to a substantial loss of MeCP2 from striatum although small numbers of MeCP2-expressing cells remain. Scale bar corresponds to 100 μm.

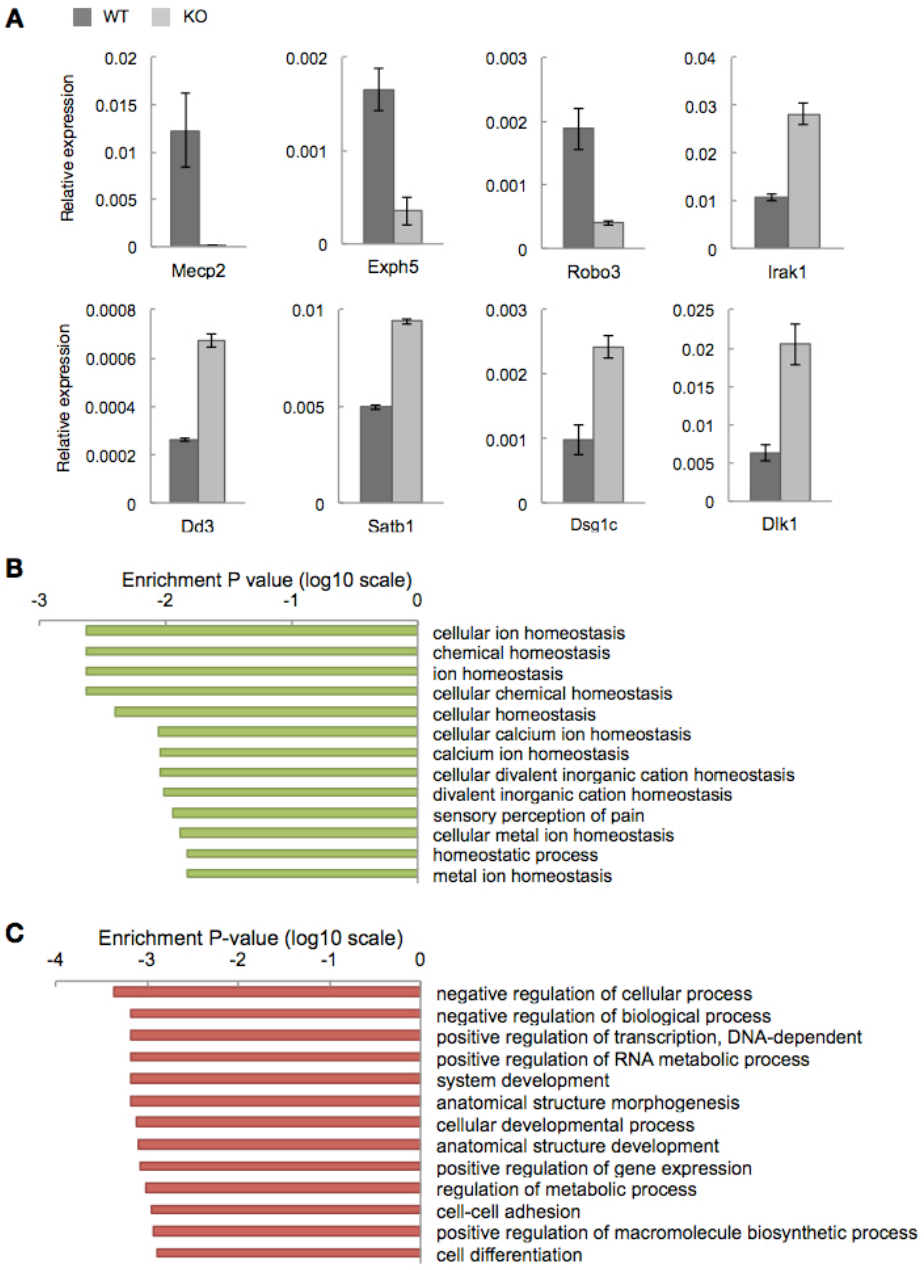

FIGURE 2.3: Gene expression changes in the striatum of *Mecp2* constitutive knockout mice

(A) Validation of differentially expressed genes in striatum of WT and constitutive knockout (*Mecp2*-null) mice by quantitative real-time RT-PCR.

(B) Gene ontology (GO) analysis for genes enriched in the striatum of *Mecp2*-null mice compared to their WT littermates.

(C) GO analysis for genes depleted in the striatum of *Mecp2*-null mice compared to their WT littermates. The enrichment p-values were calculated by hypergeometric test and adjusted by Yekutieli method using GOEAST software.

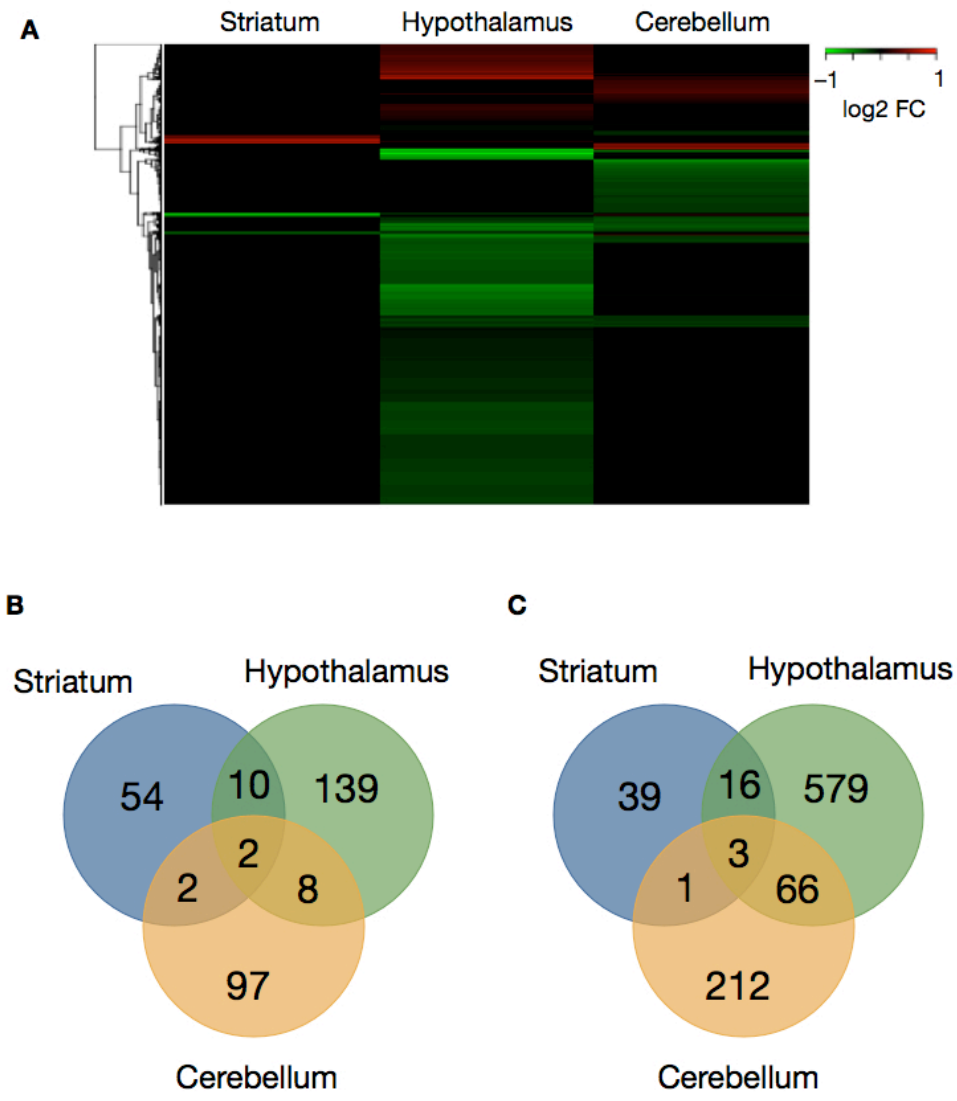

FIGURE 2.4: Differential gene expression in different brain regions of *Mecp2*-null mice

(A) Heat map of the differentially expressed genes in striatum, hypothalamus and cerebellum of *Mecp2*-null mice. Expression values are color-coded according to the legend to the right. The dendrogram depicts unsupervised hierarchical clustering based on fold-changes.

(B) Venn diagram depicting the overlap between genes differentially up-regulated in striatum, hypothalamus and cerebellum.

(C) Venn diagram depicting the overlap between genes differentially down-regulated in striatum, hypothalamus and cerebellum.

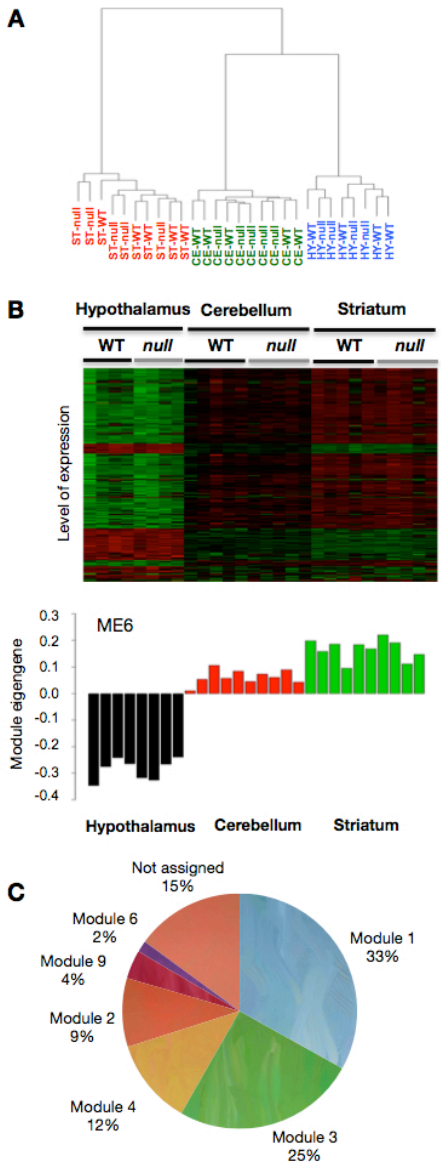

FIGURE 2.5: Gene expression network analysis in different brain regions of WT and *Mecp2*-null mice

(A) Unsupervised hierarchical clustering of the gene expression profiles of striatum (ST), hypothalamus (HY), and cerebellum (CE) in *Mecp2*-null (null) and WT mice.

(B, top) Heat map of the gene expression profiles in *Mecp2*-null and WT data sets across three brain regions. **(B, bottom)** A representative module (ME6) of co-regulated genes generated based on topological overlap of gene expression data from WT and KO mice. A clear separation of eigengene values can be seen based on brain region but not genotype.

(C) Localization of differentially expressed genes within modules assigned using WGCNA in the striatum.

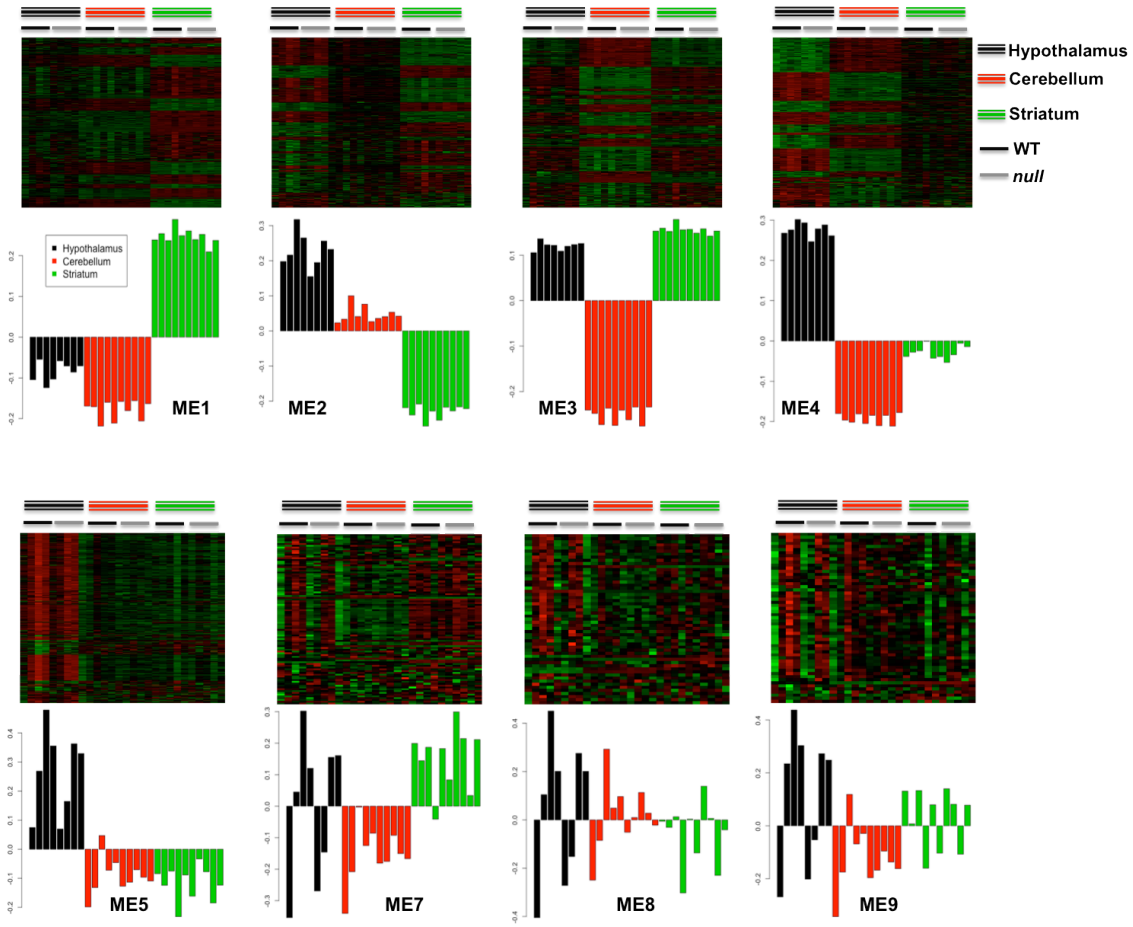

FIGURE 2.6: Weighted gene co-expression network analysis (WGCNA) in *Mecp2*-null and wild-type mice

Gene expression profiles and module eigengene of the 8 modules in the co-expression network constructed using the *Mecp2*-null and wild-type data sets.

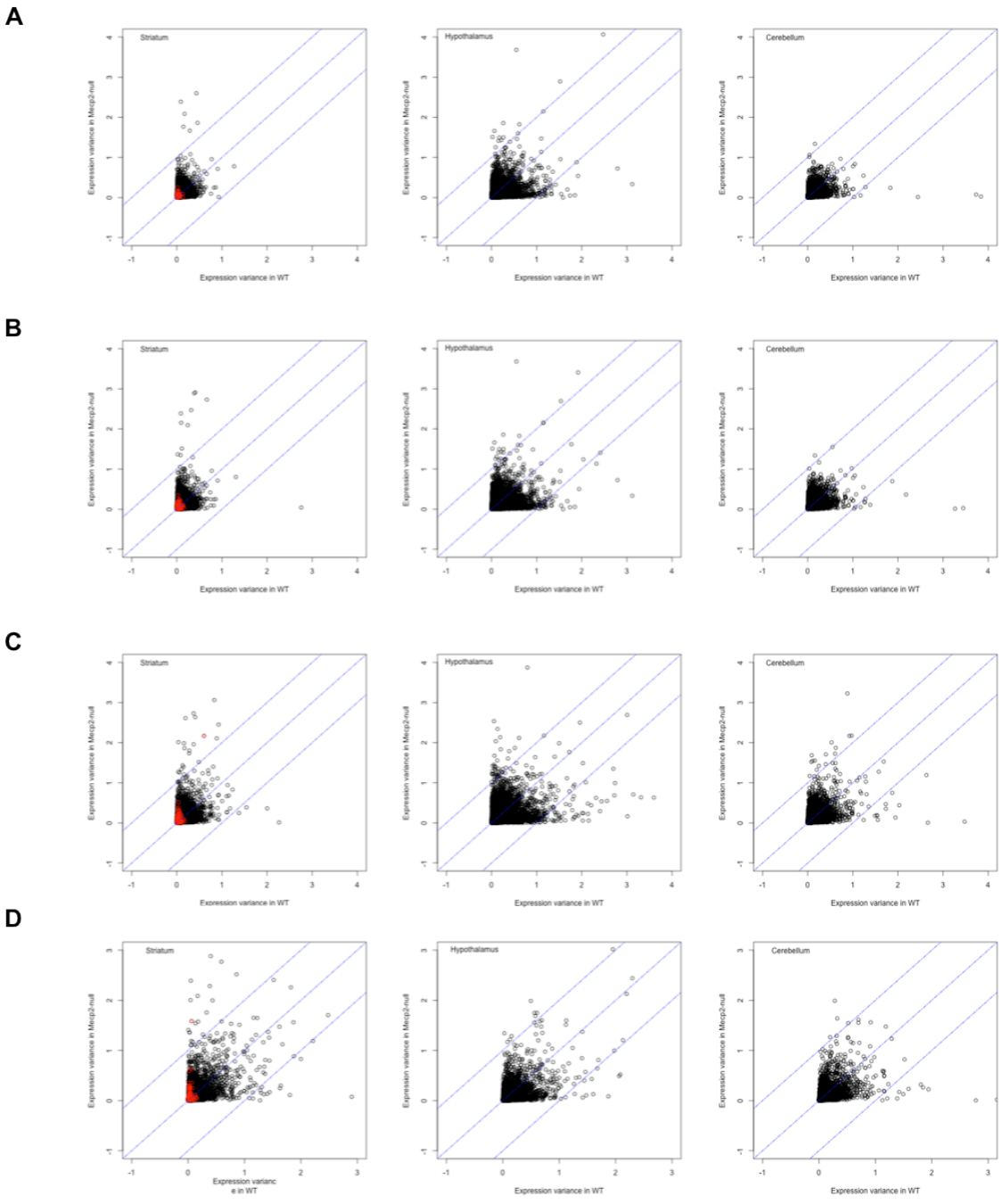

FIGURE 2.7: Gene expression variances between replicates in striatum, hypothalamus, and cerebellum in *Mecp2*-null and wild type (WT) mice

Gene level expression profiles used in this analysis were generated by the mean **(A)**, median **(B)**, APT **(C)**, and JETTA **(D)** methods, respectively. Differentially expressed genes between *Mecp2*-null and WT mice in striatum are labeled in red.

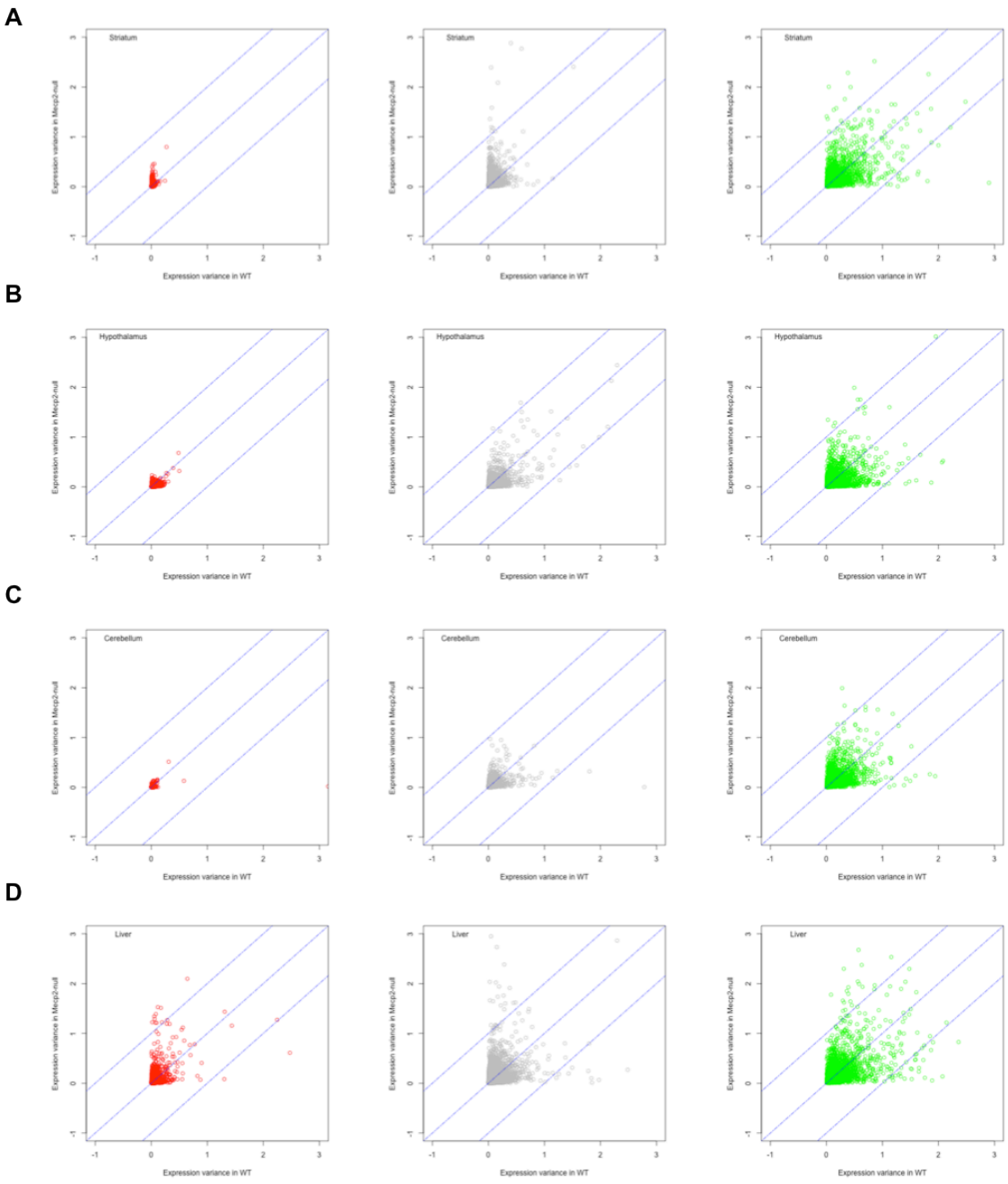

FIGURE 2.8

Expression variances between replicates for the highly-expressed (red), intermediately-expressed (grey), and lowly-expressed (green) genes in striatum **(A)**, hypothalamus **(B)**, cerebellum **(C)**, and liver **(D)** in *Mecp2*-null and wild type (WT) mice. Gene level expression profiles used in this analysis were generated by the JETTA package.

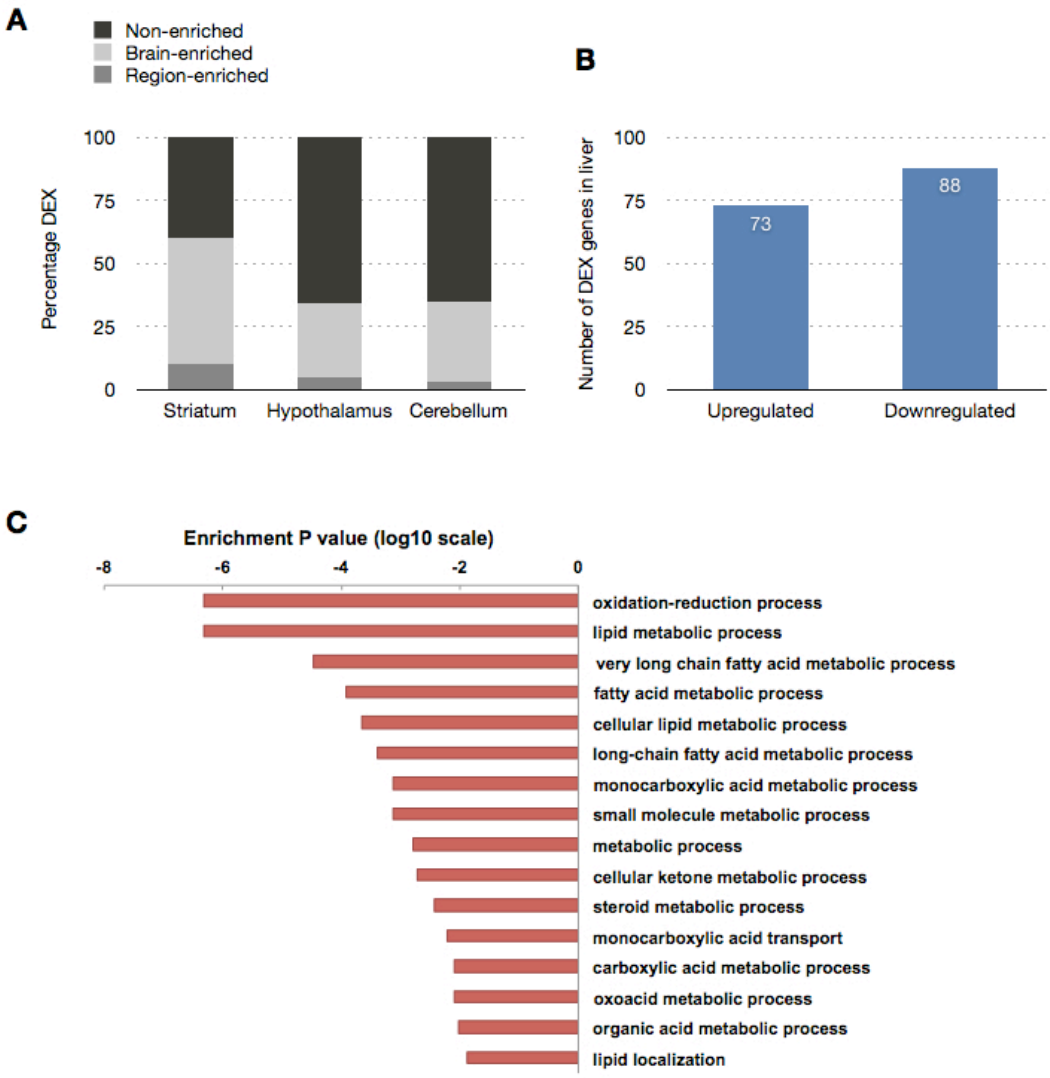

FIGURE 2.9 Distribution of differentially expressed genes in WT and *Mecp2*-null mice

(A) Genes differentially expressed in WT and KO tissues were separated into those that show enrichment within their representative brain region, those enriched in brain versus liver, and those that are not enriched.

(B) The number of genes up-regulated and down-regulated in liver in *Mecp2*-null mice compared with WT littermates.

(C) Gene ontology (GO) analysis showing biological functions of the genes up-regulated in the liver of *Mecp2*-null mice compared to their WT littermates. The enrichment p-values are calculated by hypergeometric test and adjusted by Yekutieli method using GOEAST software.

CHAPTER 3: Biotin tagging of MeCP2 reveals contextual insights into the Rett syndrome transcriptome

Adapted from: Johnson, B.S.*, Zhao, Y.*, Fasolino, M.D.*, Lamonica, J., Kim, Y.J., Wood, K.H., Bu, D., Cui, Y., Goffin, D., Kim, T.H., Zhou, Z. Biotin tagging of MeCP2 reveals contextual insights into the Rett syndrome transcriptome. *Manuscript in preparation.*

ABSTRACT

Mutations in *MECP2* are responsible for Rett syndrome (RTT), a severe X-linked neurological disorder characterized by loss of developmental milestones, intellectual disability and motor impairments. Molecular insight into how these mutations disrupt neuronal function and contribute to RTT is impeded by the cellular heterogeneity of the mammalian brain. We generated knock-in mice bearing frequent RTT mutations (T158M, R106W) in conjunction with cell type-specific biotin tagging and profiled the transcriptome of WT and mutant neurons across two distinct cell types and ages. We found that MeCP2-dependent gene expression varies by age and cell type, and the degree of transcriptional changes correlates with the severity of the RTT mutation. We also provide evidence supporting post-transcriptional compensation of misregulated long genes. Finally, we circumvent genetic mosaicism associated with female mouse models of RTT and identify functionally distinct gene expression changes between neighboring WT and mutant neurons, therefore providing key insights into RTT.

INTRODUCTION

RTT is a progressive X-linked neurological disorder that represents one the most common causes of mental disability among young girls, occurring in 1 out of every 10,000 female births. Individuals with classic RTT appear asymptomatic until 6-18 months of age, after which they experience a characteristic loss of acquired language, social, and psychomotor skills, followed by the manifestation of stereotyped hand movements, hypotonia, breathing irregularities, and, in a large majority of patients, seizures (Neul et al., 2010; Percy et al., 2010). Approximately 95% of RTT cases are genetically mapped to the X-linked gene that encodes Methyl-CpG binding protein 2 (MeCP2), a ubiquitously expressed protein that is highly enriched in post-mitotic neurons of the mammalian brain relative to other somatic cell types (Cuddapah et al., 2014; Shahbazian et al., 2002b). The majority of RTT missense and nonsense mutations associated with *MECP2* are predominantly clustered in two functionally distinct protein domains that are critical for MeCP2 function. The Methyl-CpG Binding Domain (MBD) allows MeCP2 to recognize and bind to symmetrically methylated cytosines *(Lewis et al., 1992; Nan et al., 1997)*. The Transcriptional Repression Domain (TRD) mediates protein-protein interactions with histone deacetylase-containing co-repressor complexes, including the NCoR-SMRT and mSin3A complexes (Ebert et al., 2013; Lyst et al., 2013; Nan et al., 1998). Together these domains implicate MeCP2 as a nuclear protein that is broadly distributed throughout chromatin and likely mediates transcriptional repression (Nan et al., 1997; Skene et al., 2010), although transcriptional activation has also been reported (Chahrour et al., 2008; Li et al., 2013a).

Although much research into RTT has primarily focused on hemizygous male mice, due to the limits imposed by random X-inactivation in female mice, animal models bearing mutations in *Mecp2* faithfully recapitulate the pathophysiology of human RTT patients. Following apparently normal post-natal development, *Mecp2*-null male mice progressively display abnormal gait and hypoactivity, breathing irregularities, learning and memory impairments, and premature lethality, which are accompanied by morphologial deficits and electrophysiological abnormalities in the brain (Chen et al., 2001; Guy et al., 2001). *Mecp2*-mutant mice bearing RTT missense mutations

further model the clinical variability associated with RTT, highlighting the individual molecular contributions of the MBD and TRD towards RTT phenotypes. However, despite clear genetic linkage and conserved pathophysiology, a mechanism by which MeCP2 loss of function contributes to the cellular, synaptic, circuit, and behavioral deficits seen in RTT patients and *Mecp2* mouse models remains poorly understood.

The identification of transcriptional targets associated with MeCP2 dysfunction remains a key step towards illuminating the molecular etiology of RTT. Previous studies have corroborated that subtle gene expression changes occur among numerous genes that are misregulated in *Mecp2*-mutant male mice and in cultured neurons lacking MeCP2. However, the phenotypic consequences of such effects are unclear, having only been explored on a gene-specific basis despite lists of misregulated genes that vary across multiple studies (Ben-Shachar et al., 2009; Chahrour et al., 2008; Jordan et al., 2007; Kriaucionis et al., 2006; Li et al., 2013a; Nuber et al., 2005; Sugino et al., 2014; Tudor et al., 2002; Urdinguio et al., 2008). Reports detailing MeCP2 binding to methylated cytosines (mCpG and mCpA), hydroxymethylated cytosines (hmCpG), and GC-rich regions throughout the neuronal genome also remain conflicting, thus making the identification of direct MeCP2 target genes difficult to ascertain (Chen et al., 2015; Cohen et al., 2011; Gabel et al., 2015; Rube et al., 2016; Skene et al., 2010).

The complexity of studying MeCP2 transcriptional function is further confounded by the heterogeneity of cellular transcriptomes and their diversity in the mammalian brain. The brain comprises a myriad of intermixed cell types that differ in morphology, function, and electrophysiological properties. Current studies are now beginning to uncover the distinctive transcriptional and epigenomic programs that vary between different cell types in order to maintain this cellular diversity (Fuzik et al., 2016; Mo et al., 2015; Tasic et al., 2016; Usoskin et al., 2015). Genomic distributions of methylated and hydroxymethylated cytosines are cell type-specific, but our knowledge remains limited on whether MeCP2, despite its ubiquitous expression throughout the brain, exerts transcriptional control in a cell type-dependent or independent manner (Sugino et al., 2014). Addressing this heterogeneity and investigating transcriptional

changes within the varying contexts of individual cell types may provide key insights into how MeCP2 modulates gene expression.

Within individual cell types, RNA also exists as a heterogeneous pool of nuclear and cytoplasmic transcripts at different stages of synthesis, processing and maturation, sub-cellular localization, translation, and degradation. The bulk of protein-coding mRNAs are enriched in the cytoplasm and subject to post-transcriptional regulation. In contrast, protein-coding RNAs in the nucleus represent transcripts not yet heavily influenced by post-transcriptional regulation, and are thus well suited for studying transcriptional dynamics in the cell (Bhatt et al., 2012; Djebali et al., 2012). Measured gene expression changes using whole cell mRNA from individual brain regions thus reflect composite profiles that easily obscure cell type-specific expression changes, which may also impede an accurate assessment of MeCP2 function at the transcriptional level.

In this study, we addressed these issues by engineering a collection of genetically modified mice whereby nuclear MeCP2 is “tagged” with biotin in a Cre-recombinase dependent manner. To understand the molecular impact of RTT-associated mutations on cell type-specific gene expression *in vivo*, we also developed tagged knockin mice bearing T158M or R106W RTT missense mutations. When combined with Fluorescence-Activated Cell Sorting (FACS), this strategy effectively circumvents the cellular heterogeneity of the mouse brain and allows isolation of nuclei from cell types of interest. By examining MeCP2 differentially expressed genes (FDR < 0.05, herein DEGs) within the context of neuronal gene expression in the nucleus, we identified transcriptional similarities and differences that correlate with the severity of the mutation in different cell types and at different ages. Furthermore, we find that genome-wide transcriptional deficits in the nucleus are balanced by post-transcriptional compensation of RNAs in a gene length-dependent manner. Finally, our approach allows us to also circumvent genetic heterogeneity associated with X-inactivation in heterozygous female mice and profile transcriptional changes among neighboring WT and mutant neurons, thereby discerning cell autonomous from non-cell autonomous transcriptional effects. This comprehensive survey across different neuronal settings allows us to propose a contextualized model by which cell and non-cell

autonomous transcriptional changes in different cell types contribute to the molecular severity of neuronal deficits observed in RTT.

MATERIALS AND METHODS

Generation and Phenotypic Characterization of Mice

The targeting construct used for homologous recombination at the *Mecp2* locus in murine ES cells was cloned in two arms by PCR amplification of sv129 genomic DNA. The 5′ arm was PCR amplified with 5′-AGGAGGTAGGTGGCATCCTT-3′ and 5′-CGTTTGATCACCATGACCTG-3′ primers, whereas the 3′ arm was PCR amplified with 5′-GAAATGGCTTCCCAAAAAGG-3′ and 5′-AAAACGGCACCCAAAGTG-3′ primers. Restriction sites at the ends of each arm were created using nested primers for cloning into a vector containing a *loxP*-flanked neomycin cassette (Neo) and a diphtheria toxin A negative-selection cassette. QuikChange (Stratagene) insertional mutagenesis was used to generate the *Mecp2-Tavi* targeting construct by inserting the Tavi tag immediately upstream of the *Mecp2* stop codon within the 5' arm. The portion of the Tavi tag containing the biotinylation consensus sequenced flanked by 5' NaeI and 3' BspHI restriction sites was inserted through two rounds of mutagenesis (Round 1 Forward: 5'-GACCGAGAGAG TTAGCGCCGGCCTGAACGACATCTTCGAGTCATGACTTTACATAGAGCG-3', Round 1 Reverse: 5'-CGCTCTATGTAAAGTCATGACTCGAAGATGTCGTTCAGGCCGGCGCTAACTCT CTCGGTC-3'; Round 2 Forward: 5'-CTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGC ACGAATCATGACTTTACATAGAG-3', Round 2 Reverse: 5'-CTCTATGTAAAGTCATGATTCGT GCCATTCGATTTTCTGAGCCTCGAAGATGTCGTTCAG-3'), whereas the portion of the tag containing the TEV protease cleavage site was inserted upstream of the NaeI restriction site with a third round of mutagenesis (Round 3 Forward: 5'-GACCGAGAGAGTTAGCGAAAACCTG TATTTTCAGGGCGCCGGCCTGAACGACATC-3', Round 3 Reverse: 5'-GATGTCGTTCAGGCC GGCGCCCTGAAAATACAGGTTTTCGCTAACTCTCTCGGTC-3'). To generate *Mecp2-Tavi* targeting constructs bearing independent RTT-associated point mutations, QuikChange site-directed mutagenesis was used to mutate MeCP2 arginine 106 to tryptophan and MeCP2

threonine 158 to methionine within the 3'arm and 5'arm, respectively. A single nucleotide at codon T160 also underwent site-directed mutagenesis for a silent mutation to introduce a BstEII restriction site to correctly identify targeted ES cells. To generate conditional *BirA* transgenic mice, PCR primers containing AscI restriction sites and a Kozak consensus sequence were used to subclone the *BirA* coding sequence and insert it downstream of the floxed Neo-STOP cassette within pROSA26-1, a transgenic targeting vector that has previously been characterized.

After confirmation by Sanger sequencing and linearization with NotI (*Mecp2-Tavi* targeting construct and its mutant variants) or SgfI (*BirA* targeting construct), the constructs were electroporated into sv129-derived murine ES cells. Correctly targeted ES cells were independently injected into C57BL/6 blastocysts and subsequently implanted into pseudopregnant females. Agouti offspring were screened by PCR genotyping to confirm germline transmission of the *Mecp2-Tavi*, $Mecp2^{T158M}$*-Tavi*, $Mecp2^{R106W}$*-Tavi*, and $R26^{cBirA}$ alleles. In the case of the *Mecp2-Tavi* allele and its mutant variants, the resulting offspring were mated with C57BL/6 *EIIa-cre* mice to ensure germline deletion of the floxed Neo cassette between *Mecp2* exons 3 and 4.

Mouse lines

Dlx5/6-Cre mice (Stock 008199) were obtained from Jackson Laboratories. *NeuroD6/NEX-Cre* mice were obtained with permission from the Nave Laboratory.

Animal Husbandry

Experiments were conducted in accordance with the ethical guidelines of the US National Institutes of Health and with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania. All of the experiments described were performed using mice on a

congenic sv129:C57BL/6 background with the knock-in/transgenic alleles backcrossed to C57BL/6 mice (Charles River) for at least five generations, unless otherwise stated. Mice were housed in a standard 12h light/12h dark cycle with access to ample amounts of food and water. Mice bearing the Tavi tag were genotyped using a bipartite primer PCR-based strategy to detect the Tavi tag at the 3'-end of the endogenous *Mecp2* gene (Forward: 5'-CACCCCGAAGCCACGAAACTC-3', Reverse: 5'-TAAGACTCAGCCTATGGTCGCC-3') and give rise to a 318-bp product from the wild-type allele and a 388-bp product from the tagged allele. Mice bearing the *BirA* transgene were genotyped using a tripartite primer PCR-based strategy to detect the presence or absence of the CAG promoter at the *Rosa26* locus (Forward:5'-TGCTGCCTCCTGGCTTCTGAG-3', Reverse #1: 5'-GGCGTACTTGGCATATGATACAC-3', Reverse #2: 5'-CACCTGTTCAATTCCCCTGCAG-3') and give rise to a 173-bp product from the wild-type allele and a 477-bp product from the transgene-bearing allele. Mice bearing Cre-recombinase (either *NeuroD6/NEX-Cre* or *Dlx5/6-Cre*) were genotyped using PCR-based strategies as previously described (REFERENCES).

Phenotypic Scoring

Phenotypic scoring was performed on a weekly basis for the absence or presence of RTT-like symptoms as described previously.

Immunofluorescence and microscopy

Mice were anesthetized with 1.25% Avertin (wt/vol), transcardially perfused with 4% paraformaldehyde (wt/vol) in 0.1M sodium-potassium phosphate buffered saline and postfixed overnight at 4°C. Brains were coronally or sagittally sectioned at 20μm using a Leica CM3050 S cryostat. Immunofluorescence on free-floating sections was performed as previously described, except sections were permeabilized with 0.5% Triton without methanol for 20 minutes, and

sections were blocked overnight with 10% Normal Goat Serum and 1:100 unconjugated goat anti-mouse IgG (Sigma M5899). The following primary antibodies were incubated at 4°C overnight: rabbit anti-MeCP2 C-terminus (1:1000, in house), rabbit anti-nucleolin (1:1000, Abcam ab22758), mouse anti-parvalbumin (1:500, Millipore MAB1572), rabbit anti-calretinin (1:1000, Swant 7699/3H), mouse anti-GAD67 (1:500, Millipore MAB5406), mouse anti-NeuN (1:500, Millipore MAB377). For rat anti-somatostatin (1:250, Millipore MAB354MI), primary incubation was performed for 48 hours at 4°C. Flourescence detection of primary antibodies was performed using Alexa 488-conjugated goat anti-rabbit (1:1000, Invitrogen A11008), Alexa 488-conjugated goat anti-mouse (1:1000, Invitrogen A11029), and Alexa 488 goat anti-rat (1:1000, Invitrogen A11006). Fluorescence detection of biotin was performed simultaneously with secondary antibody incubations, using Streptavidin Dylight 650 (1:1000, Fisher 84547) for fluorescence microscopy and Streptavidin Dylight 550 (1:1000, Fisher 84542) for confocal microscopy. Sections were counterstained with DAPI (1:1000, Affymetrix 14564) to visualize DNA before mounting with Fluoromount G (SouthernBiotech). Images were acquired using a Leica DM5500B fluorescent microscope with a Leica DFC360 FX digital camera (region-specific biotinylation, quantification of neuronal cell type-specific markers) or a Leica TCS SP8 Multiphoton confocal microscope (representative images of neuronal cell type specific markers, subcellular localization of MeCP2). Images were acquired using identical settings for laser power, detector gain amplifier offset and pinhole diameter in each channel. Image processing was performed using ImageJ and Adobe Photoshop, including identical adjustments of brightness, contrast, and levels in individual color channels for merged images across genotypes.

Quantitative western analysis

Quantitative western blot was performed using Odyssey Infrared Imaging System (Licor). Primary antibodies include rabbit anti-MeCP2 C-terminus (1:4000, in house), mouse anti-MeCP2 N-terminus (1:4000, Sigma M7433), mouse anti-NeuN (1:500, Millipore MAB377), and rabbit anti-

Histone H3 (1:1000, Abcam ab1791). Secondary antibodies include anti-rabbit IRDye 680LT (1:10,000, Licor), anti-mouse IRDye 800CW (Licor), Streptavidin Dylight 650 (1:10,000, Fisher 84547) and Streptavidin Dylight 800 (1:10,000, Fisher 21851). Quantification of protein expression levels was carried out following Odyssey Infrared Imaging System protocols.

Co-immunoprecipitation using nuclear extracts

Tissues were mined on ice and homogenized in ice cold lysis buffer (10 mM HEPES pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5% NP-40, 0.2mM EDTA, protease inhibitors). Nuclei were pelleted, washed and resuspended in nuclear extract (NE) buffer (20mM HEPES pH 7.9, 1.5mM $MgCl_2$, 500mM KCl, 0.2mM EDTA, 10% glycerol, protease inhibitors). Nuclei were incubated in NE buffer at 4°C for two hours with rotation. Samples were cleared by ultracentrifugation with a TLA 100.3 rotor (Beckman Optima TL) at 4°C for 30 minutes and the supernatant taken for nuclear extract. Protein concentration was quantified using a modified Bradford assay (Bio-Rad). 1mg of nuclear extract was adjusted to 300µl total volume with NE buffer to perform IP in duplicate. Protein G Dynabeads or Streptavidin M-280 Dynabeads (Life Technologies) were washed three times in PBS with 0.1% Tween-20 and 0.1% BSA. Nuclear extracts were cleared for 30 minutes at 4°C with 25µl Protein G Dynabeads. For streptavidin pulldown, 50µl of Streptavidin M-280 Dynabeads were added to the nuclear extract and incubated at 4°C for two hours with rotation. To test if the Tavi tag was required for streptavidin pulldown, nuclear extracts were split and incubated with or without 200U TEV protease (Invitrogen) in the absence of a reducing agent and without agitation at 4°C for ≥ 4 hours prior to IP. For antibody immunopreciptiation, 5µg of (rabbit anti-MeCP2, in house) was added to the nuclear extract and incubated overnight at 4°C with rotation. Protein G beads were blocked in wash buffer overnight at 4°C with rotation. Blocked beads were then incubated with antibody-bound nuclear extract for two hours at 4°C with rotation. Beads were washed four times in PBS with 0.1% Tween-20 and split into two equal volumes. Each sample

was resusptended in 25µl loading buffer with 50mM DTT and boiled for 10 minutes at 95°C prior to loading on a 4-12% Bis-Tris NuPage gel (Life Technologies).

Chromatin immunoprecipitation

Forebrain tissues from male mice at 8 weeks were homogenized in cross-linking buffer (1% formaldehyde (wt/vol), 10mM HEPES (pH 7.5), 100mM NaCl, 1mM EDTA, 1mM EGTA) and cross-linked for 5 minutes at RT. After quenching with 125mM glycine, cross-linked tissue was washed with ice-cold PBS and dounced with 16 strokes in lysis buffer (50mM HEPES (pH 7.5), 140mM NaCl, 1mM EDTA, 1mM EGTA, 10% glycerol (vol/vol), 0.5% NP-40 (vol/vol), and 0.25% Triton X-100 (vol/vol) with protease inhibitors). Nulcei were pelleted, washed and resuspended in chromatin buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, and 0.5mM EGTA with protease inhibitors). Chromatin was sonicated using a Diagenode Bioruptor, and salt and detergent were added to adjust the chromatin buffer to 0.5% Triton X-100, 150mM NaCl, 10mM EDTA, and 0.1% sodium deoxycholate (DOC, vol/vol), and precleared at 4°C with Protein A Dynabeads (Invitrogen). For immunoprecipitation, 3µg of purified rabbit anti-MeCP2 IgG (in house) or non-specific rabbit IgG control (Millipore NI01) was incubated with 45µg of chromatin for 4 hours, followed by an overnight incubation with pre-blocked Protein A Dynabeads, at 4°C with rotation. Bead-bound chromatin was washed with low salt buffer (50mM HEPES pH 7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC), high salt buffer (50mM HEPES pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC), LiCl buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA, 0.5% NP-40, 0.5% DOC) and TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). Chromatin was eluted with elution buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, and 1% SDS (wt/vol)), digested with proteinase K (0.5mg ml^{-1}), and reversed crosslinked at 65°C overnight. After RNase A treatment, DNA fragments were extracted with phenol/chloroform and ethanol-precipitated.

Quantitative real-time PCR (qPCR) analysis was carried out using SYBR green detection (Life Technologies) on an ABI Prism 7900HT Real-Time PCR System (Applied Biosystems). The percent input for each amplicon was determined by comparing the average threshold cycle of the immunoprecipitated DNA to a standard curve generated using serial dilutions of the input DNA and interpolating the “fraction of input” value for this sample.

RT-PCR and RNA-seq using RNA from FACS-isolated nuclei

Total RNA was prepared from FACS-isolated cortical nuclei of male mice at 6 weeks (TAVI, T158M, R106W, 2-3 mice pooled per biological replicate, 4 replicates each), male mice at 12 weeks (TAVI, T158M, 2 mice pooled per biological replicate, 3 replicates each), and female mice at 18 weeks (TAVI, T158M, R106W, single mouse per biological replicate, 2 replicates each). Nuclei were isolated from fresh cortical tissue for FACS as previously described under ice-cold and nuclease-free conditions. Mouse cortices were rapidly resected on ice and subjected to dounce homogenization in homogenization buffer (0.32M sucrose, 5mM CaCl2, 3mM MgAc2, 10mM Tris-HCl pH 8.0, 0.1% Triton, 0.1mM EDTA, Roche Complete Protease Inhibitor without EDTA). Homogenates were layered onto a sucrose cushion (1.8M sucrose, 10mM Tris-HCl pH 8.0, 3mM MgAc2 Roche Complete Protease Inhibitor without EDTA) and centrifuged at 25,000 rpm at 4°C for 2.5 hours using a Beckman SW25.1 swinging bucket rotor. Nuclei were resuspended & washed once in blocking buffer (1x PBS, 0.5% BSA (Sigma A4503), RNasin Plus RNase Inhibitor (Promega)) and pelleted using a tabletop centrifuge at 5000 RCF at 4°C for 10 minutes. Nuclei were resuspended in blocking buffer to a concentration of ~$6x10^6$ nuclei/ml, blocked for 20 minutes at 4°C with rotation, then incubated with Streptavidin Dylight 650 (1:1000, Fisher 84547) and Alexa 488-conjugated anti-NeuN antibody (1:1000, Millipore MAB377X) for 30 minutes at 4°C with rotation. After a 5-minute incubation with 1:1000 DAPI to enable singlet detection during FACS, labeled nuclei were washed for an additional 30 minutes at 4°C with blocking buffer, pelleted and resuspended in blocking buffer with 1% BSA. A BD Biosciences

Influx cell sorter at the University of Pennsylvania Flow Cytometry and Cell Sorting Facility was used to identify cell type-specific populations of nuclei, and 1.2 – 2.5 x10^5 singlet nuclei from specified populations were directly sorted into Qiagen Buffer RLT Plus for immediate lysis and stabilization of RNA transcripts. Total nuclear RNA was processed using the Qiagen AllPrep DNA/RNA mini kit according to manufacturer instructions, with exception to the on-column DNaseI treatment. RNA was eluted from RNeasy mini spin columns and treated with DNaseI (Qiagen 79254) for 25 minutes at room temperature, then precipitated with glycogen/NaOAc and stored in ethonal at -80°C. Ethanol precipitation of nuclear RNA was carried out to completion prior to initiating RT-PCR or RNA-seq library construction.

To validate cell type-specific cortical nuclei populations, total RNA from a single aliquot of sorted nuclei was converted to cDNA with random hexamers using the SuperScript III First-Strand Synthesis System (Invitrogen). RT-PCR was performed on a ABI Prism 7900HT Real-Time PCR System (Applied Biosystems) using exon-spanning Taqman gene expression assays to detect mRNA transcripts for the following genes: *CRE* (Mr00635245_cn), *Mecp2* (Mm01193537_g1), *Rbfox3* (Mm01248771_m1), *Gfap* (Mm01253033_m1), *Aif1* (Mm00479862_g1), *Mog* (Mm00447824_m1), *Slc17a7* (Mm00812886_m1), *Tbr1* (Mm00493433_m1), *Gad1* (Mm04207432_g1), *Slc35a1* (Mm00494138_m1), *Ht3ar* (Mm00442874_m1), *Pvalb* (Mm00443100_m1), *Sst* (Mm00436671_m1), *Pgk1* (Mm00435617_m1), *Actb* (Mm00607939_s1), *ß2m* (Mm00437762_m1). A geometric mean was calculated to normalize mRNA expression levels to multiple housekeeping genes (*Actb*, *ß2m*, and *Pgk1*), and cell type-enrichment for each sorted population was determined relative to the total mixed population of DAPI+ nuclei.

For RNA-seq, the total amount of RNA isolated from a single aliquot of sorted nuclei was used as input for library construction; hence differential gene expression comparisons between FACS-isolated *Mecp2* control and mutant neurons is performed using RNA from equivalent numbers of neuronal nuclei. Total RNA was depleted of ribosomal RNAs, subjected to 5 minutes of heat fragmentation, and converted to strand-specific cDNA libraries using the TruSeq Total

RNA library prep kit with RiboZero depletion (Illumina). Multiplexed libraries were sequenced at the University of Pennsylvania Next Generation Sequencing Core facility using an Illumina HiSeq 2000/2500 that typically yielded 30-40M total reads per library.

GRO-seq

Nuclei were isolated from fresh cortical tissue under ice-cold and nuclease-free conditions as described in the preceding section. After ultracentrifugation, nuclei were resuspended & washed once in PBS (1x PBS, RNasin Plus RNase Inhibitor (Promega)) and pelleted using a tabletop centrifuge at 5000 RCF at 4°C for 10 minutes. Nuclei were resuspended in PBS, pipetted through a 0.22μm filter and counted using a hemocytometer. Nuclei were then pelleted, resuspended to a concentration of 5×10^6 - 10×10^6 nuclei/100μl in glycerol storage buffer (50mM Tris pH 8.3, 40% glycerol, 5mM $MgCl_2$, 0.1 mM), and flash frozen in liquid N_2 for storage until needed.

For each nuclear run-on (NRO), 100μl of nuclei was mixed with 46.5μl NRO Reaction Buffer (10mM Tris pH 8.0, 5mM $MgCl_2$, 1 mM DTT, 300mM KCl), 3.5μl Nucleoside Mix (50μM ATP, 50μM GTP, 2μM CTP, 50μM Br-UTP, 0.4U/μl RNasin), and 50μl 2% Sarkosyl. The NRO reaction was performed at 30°C for 5 minutes, then terminated by a 20 minute incubation with DNAse I at 37°C, followed by a hour-long incubation with 225μl NRO Stop Buffer (20mM Tris, pH 7.4, 10mM EDTA, 2% SDS) and Proteinase K at 55°C. Phenol-extracted RNA was fragmented with 0.2N NaOH, and BrU-RNA was isolated three consecutive times with BrdU-antibody beads, with enzymatic TAP and PNK treatments to remove the cap and 3'-phosphate and to add a 5'-phosphate, as well as Illumina TruSeq small RNA sample prep kit adapter ligations between BrU-RNA isolation steps as described in Core et al. 2008, Greer et al. 2016.

RNA-seq mapping and analysis

The FASTQ files of each samples were mapped to the mouse mm10 genome by STAR (version 2.3.0) using the parameters of "--outFilterMultimapNmax 1 --outFilterMismatchNmax 3". The number of reads mapped to the whole gene body of each gene were counted using Perl scripts generated in-house. All comparisons were performed by the edgeR and DEseq2 R packages, and FDR < 0.05 was the criteria used to identify differentially expressed genes. For each comparison, the results from edgeR and DESeq2 were merged to avoid method-based biases. The mean fold changes and the mean FDRs generated from both methods were used for generating plots and heatmaps.

Statistical Analyses

Statistical analyses were performed using Prism 6.0 (GraphPad Software) and R. Individual statistical tests are stated in the figure legends.

RESULTS

Engineering a System to Genetically Biotinylate MeCP2 *In Vivo*

Biotin-mediated affinity tagging has been utilized in multiple cell and animal models for genomic, biochemical, and imaging experimental approaches due to the strong and highly specific interaction between biotin and streptavidin protein (K_d = 4x10^{-14}M) (Boer et al., 2003; Driegen et al., 2005; Kim et al., 2009; Malik and Henikoff, 2010; Ooi et al., 2010). To investigate MeCP2 function within Cre-specified cell types, we engineered a short 23-amino acid Tavi (*T*EV and *Avi*din-binding) affinity tag (Figure 3.1A). This tag comprises a 15-amino acid consensus motif that is posttranslationally labeled with biotin in the presence of the *E. coli* biotin ligase BirA, as well as a TEV protease cleavage site. We used homologous recombination to insert the Tavi tag immediately upstream of the stop codon at the *Mecp2* genomic locus, thereby generating mice where endogenous MeCP2 is biotinylated.

To biotinylate the tag in a conditional manner, we cloned the BirA gene into a separate targeting construct. A CAG promoter and a floxed *Neo-STOP* cassete, which acts as a strong attenuator of transcription, preceded the BirA gene, followed by an IRES sequence and a single copy of GFP (*CAG-loxP-Neo-STOP-loxP-BirA-IRES-GFP*) (Figure 3.2A). This transgene was targeted to the *Rosa26* gene locus to generate conditional BirA mice ($R26^{STOP\text{-}BirA}$). In a mouse that expresses Cre recombinase within a genetically defined cell population, the *STOP* cassette is removed, allowing BirA to be expressed and MeCP2-Tavi to be biotinylated (Figure 3.1B). We tested this by using *EIIa-Cre* to excise the *STOP* cassette from the germline, allowing BirA to be ubiquitously expressed throughout the body ($R26^{BirA}$). Notably, $R26^{BirA}$ heterozygous and homozygous mice remained viable, fertile, and devoid of gross abnormalities, consistent with previously engineered transgenic mice that express BirA either ubiquitously or within restricted tissues using cell type-specific promoters (Boer et al., 2003; Driegen et al., 2005). We crossed $R26^{STOP\text{-}BirA/+}$ and $R26^{BirA/+}$ males to $Mecp2^{Tavi/+}$ female mice to produce male progeny and assess the specificity of BirA expression and enzymatic activity. Whole brain nuclear extracts revealed that MeCP2-Tavi is detected by streptavidin, and thus successfully biotinylated, only under conditions where BirA is expressed ($R26^{BirA/+}$;$Mecp2^{Tavi/y}$) (Figure 3.1C). Biotinylation *in vivo* was restricted to the Tavi tag as MeCP2 streptavidin signal was absent from both $R26^{BirA/+}$;$Mecp2^{+/y}$ nuclear lysates and $R26^{BirA/+}$;$Mecp2^{Tavi/y}$ lysates pre-treated with TEV protease to remove the Tavi tag prior to immunoblotting (Figure 1C). Immunoprecipitation of MeCP2 from whole brain nuclear extracts using streptavidin-coated magnetic beads yielded similar observations (Figure 3.2B). Thus streptavidin specifically recognizes MeCP2 when it is attached to a biotinylated Tavi tag.

Normal MeCP2 function is perturbed through altered protein expression levels, reduced binding to methylated DNA, or disrupted interactions with known co-repressor complexes and leads to the appearance of RTT-like phenotypes in mice (Brown et al., 2015; Collins et al., 2004; Ebert et al., 2013; Goffin et al., 2012; Lyst et al., 2013; Samaco et al., 2008). To determine if MeCP2-Tavi remains functional *in vivo*, we assessed whole brain RNA and protein expression levels from age-matched adult $Mecp2^{Tavi/y}$ and control $Mecp2^{+/y}$ mice. Although RT-PCR detected

similar levels of *Mecp2* mRNA, MeCP2 protein expression levels in *Mecp2*$^{Tavi/y}$ were reduced compared to control mice (Figures 3.1D-F). However, immunofluorescence (IF) and confocal imaging of cortical neurons from paraformaldehyde-fixed brains show that MeCP2-Tavi is properly localized to the nucleus and concentrated at methyl-CpG-rich heterochromatic foci, similar to untagged MeCP2 (Figure 3.1H). Because MeCP2 nuclear localization is considered a proxy for MeCP2 binding to methylated DNA, it is unlikely that the Tavi tag interferes with MeCP2's ability to associate with chromatin *in vivo*. We performed MeCP2 chromatin immunoprecipitation (ChIP) from *Mecp2*$^{+/y}$ and *Mecp2*$^{Tavi/y}$ neuronal chromatin and observed similar levels of MeCP2 binding at methylated genomic regions by ChIP-qPCR regardless of whether the protein was tagged or untagged (Figure 3.1G), indicating that the reduction in MeCP2-Tavi protein expression likely affects nucleoplasmic, rather than chromatin-bound, fractions of MeCP2-Tavi. Furthermore, the Tavi tag does not disrupt protein-protein interactions because MeCP2-Tavi remains capable of interacting with known components of the Sin3a and NCoR-SMRT co-repressor complexes (Figure 3.2C).

Mecp2$^{+/y}$ and *Mecp2*$^{Tavi/y}$ mice were grossly indistinguishable in terms of brain weight, body weight, and lifespan (Figures 3.1K-N). We used a well-established scoring scheme that assesses several age-dependent phenotypes commonly observed in *Mecp2*-mutant mice. Using these criteria, the *Mecp2*$^{Tavi/y}$ mice were similar to *Mecp2*$^{+/y}$ mice and displayed no significant increase in phenotypic score over an observational period of 20 weeks. Therefore, in the absence of overt phenotypes typically associated with MeCP2 dysfunction, we conclude that *Mecp2*$^{Tavi/y}$ mice appear functionally equivalent to *Mecp2*$^{+/y}$ mice throughout the first 20 weeks of age.

MeCP2 Missense Mutations Recapitulate RTT-like Phenotypes in Mice

To investigate MeCP2 function using our genetic system, we examined missense mutations within functional domains of MeCP2. Four of the eight most frequent mutations associated with RTT are missense mutations, three of which are located in the methyl-CpG binding domain of human *MECP2* and include R106W (2.76% of RTT patients), R133C (4.24% of

RTT patients), and T158M (8.79% of RTT patients) (Christodoulou et al., 2003). Clinicians have observed that typical RTT patients carrying the R133C mutation are associated with milder RTT symptoms, including better preservation of ambulation and speech, whereas patients carrying the T158M or R106W mutation exhibit relatively more intermediate or severe symptoms, respectively(Cuddapah et al., 2014). Interestingly, the clinical severity of these mutations appears to scale with their effects on binding affinity to methylated DNA. *In vitro* studies have shown that MeCP2 R133C binding to methylated CpGs is indistinguishable from WT MeCP2; in contrast, MeCP2 T158M retains partial affinity, and MeCP2 R106W completely abolishes interactions between the MBD and methylated CpGs (Ballestar et al., 2000; Brown et al., 2015; Ghosh et al., 2008; Ho et al., 2008). We thus hypothesized that MeCP2-dependent gene expression changes may also scale with a partial or complete loss of binding affinity to chromatin *in vivo* and generated *Mecp2*$^{T158M\text{-}Tavi/y}$ and *Mecp2*$^{R106W\text{-}Tavi/y}$ mutant mice (Figure 3.1A).

To exclude the possibility that genetic manipulation of the *Mecp2* locus could affect MeCP2 expression when comparing between *Mecp2*$^{Tavi/y}$ (herein TAVI), *Mecp2*$^{T158M\text{-}Tavi/y}$ (herein T158M), and *Mecp2*$^{R106W\text{-}Tavi/y}$ (herein R106W) mice, we engineered these mutations using the same knock-in targeting construct that was used to create the control TAVI allele. Around 6 weeks of age when RTT-like symptoms begin to appear, mutant mice display significant decreases in MeCP2 protein expression levels despite similar levels of mRNA expression when compared to TAVI controls (Figures 3.1I-J, and 3.2D). Reduced protein stability and expression have been similarly reported in other mouse models bearing RTT missense mutations in the MBD, which include two independent mutant alleles of the T158 residue (Brown et al., 2015; Goffin et al., 2012). Furthermore, relative expression of mutant MeCP2 protein declined over early postnatal development starting from postnatal day 0 (P0), although both mutations displayed different expression kinetics over time (Figures 3.1E-F). We found that the expression of MeCP2 R106W is already reduced at P0 by ~50% and gradually declines with age. In contrast, MeCP2 T158M is indistinguishable from MeCP2 TAVI at P0, but demonstrates a more variable decline in protein expression over the same timeframe. However, MeCP2 T158M expression

levels remained consistently higher than MeCP2 R106W expression levels at each age that was examined.

To confirm the cellular localization of these mutant proteins, we performed IF and confocal imaging on BirA-expressing *Mecp2* mutant mice using a MeCP2 antibody and streptavidin. By antibody staining, MeCP2 T158M and MeCP2 R106W were diffuse throughout the nucleus and not properly localized to heterochromatic foci, which suggests that these missense mutations reduce the ability of MeCP2 to bind methyl-CpGs *in vivo*. However, streptavidin IF signal intensity was significantly reduced in T158M and R106W mice relative to TAVI mice, consistent with reduced protein instability and requiring higher sensitivity of fluorescence detection (Figure 3.2G). Although streptavidin staining confirmed a loss of MeCP2 localization to heterochromatic foci, this channel revealed a portion of mutant MeCP2 that is redistributed to the nucleolus (Figure 3.2H), similar to a previous report that found exogenous expression of GFP-tagged MeCP2 to be redistributed to the nucleolus upon deletion of the MBD and concomitant loss of binding to chromatin (Kumar et al., 2008a). We attribute the discrepant MeCP2 IF signal to the relatively small size of the streptavidin molecule (~50kDa) compared to a ~150kDa antibody, which is less affected by steric hindrance of masked epitopes.

In addition to reduced protein expression and loss of proper sub-nuclear localization, T158M and R106W mice exhibited RTT-like phenotypes that are also seen in *Mecp2*-null mice (KO), including decreased brain weight, decreased body weight, and an age-dependent increase in phenotypic score (Figures 3.1K-N). Although lifespan was significantly reduced in all three *Mecp2* mutant mice relative to WT or TAVI control mice, the median survival of R106W mice more closely resembled that of MeCP2 KO mice than T158M mice. Statistical analysis revealed a significant difference in the survival curves of T158M (median survival = 14 weeks) and R106W (median survival = 10 weeks) mice (Figure 3.1N). Taken together, we find that mutations within the same methyl-CpG binding domain can differentially affect the severity of the RTT phenotype in mice at a molecular and physiological level. Given that MeCP2 T158M is known to retain partial affinity for methyl-CpGs *in vitro*, and that our T158M mice exhibit significantly higher MeCP2

protein levels and longer survivability when compared to R106W mice, we infer that the T158M missense mutation represents a partial loss-of-function allele. In contrast, the R106W missense mutation, which abolishes affinity for methyl-CpGs *in vitro* and in mice exhibits lower MeCP2 protein levels and lower survivability on par with *Mecp2*-null mice, may represent a complete loss-of-function allele.

Genetic Biotinylation Permits Cell Type-specific Transcriptional Profiling

Given that MeCP2 is an abundantly expressed nuclear protein present in all neurons of the CNS (Shahbazian et al., 2002b; Skene et al., 2010), we utilized our genetic biotinylation system for cell type-specific transcriptional profiling (Figure 3.3B). We used the *NeuroD6/NEX-Cre* line (Goebbels et al., 2006) to drive BirA expression in forebrain glutamatergic (excitatory) neurons and verified that MeCP2-Tavi was biotinylated in the appropriate cell types of several brain regions using fluorescence and confocal microscopy (Figures 3.3A, 3.4A-H). IF and quantification of pan-neuronal (NeuN), pan-inhibitory (GAD67), and inhibitory-specific (parvalumin, somatostatin, calretinin) neuronal markers in the somatosensory cortex of $Mecp2^{Tavi/y}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$ (NEX-Cre) mice indicates that biotinylation occurs in approximately 80% of NeuN+ cortical neurons which are also devoid of inhibitory markers, consistent with these cells being excitatory neurons (Figure 3.4H).

Fluorescence-Activated Cell Sorting (FACS) of singlet DAPI+ cortical nuclei stained for both NeuN and biotin also revealed a similar distribution of NeuN+Biotin+ (NEX-Cre blue, 85.2% ± 0.35) and NeuN+Biotin- (NEX-Cre red, 14.8% ± 0.35) nuclei, as well as a third population that was NeuN-Biotin- (NEX-Cre gray) (Figure 3.3C). We sorted for 120,000 nuclei from each of the three populations with high purity (98-99%) and isolated total nuclear RNA to perform real-time PCR (RT-PCR) for several cell type-specific markers (Figure 3.3D). The detection of both Cre-recombinase and glutamatergic-specific transcripts (*Slc17a7, Tbr1*) within NeuN+Biotin+ nuclei confirmed that this population represents excitatory neurons. In contrast, NeuN+Biotin- nuclei, which lacked Cre-recombinase expression and were exclusively enriched for GABAergic-specific

transcripts (*Gad1*, *Slc35a1*, *Ht3ar*, *Pvalb*, *Sst*), were representative of a collection of inhibitory interneuron subtypes. Finally, both NeuN+Biotin+ and NeuN+Biotin- populations were depleted of non-neuronal astrocyte, microglia and oligodendrocyte markers (*Gfap*, *Iba1*, *Mog*, respectively), which were restricted to the NeuN-Biotin- population. We independently confirmed these observations by repeating these IF and RT-PCR experiments using *Mecp2*$^{Tavi/y}$;*R26*$^{cBirA/+}$;*Dlx5/6*$^{Cre/+}$ (Dlx5/6-Cre) mice to drive BirA expression in GABAergic inhibitory neurons and observed inverted but similar results to NEX-Cre mice (Figures 2A and S2A-J) (Potter et al., 2009). We find that despite the known limitations associated with transgenic Cre lines , MeCP2-Tavi can be precisely biotinylated in the appropriate cell types and used for cell type-specific transcriptional profiling.

Phenotypes associated with the genetic removal of MeCP2 from specific brain regions or cell types in mice are largely non-overlapping and likely reflect the neural circuit that is being impaired, thus context-specific (Chao et al., 2010; Fyffe et al., 2008; He et al., 2014; Ito-Ishida et al., 2015; Samaco et al., 2009). We therefore questioned whether MeCP2-dependent gene expression changes *in vivo* were also context-specific given the differences in gene expression between individual neuronal cell types, and whether these expression changes would be conserved across independent *Mecp2* mutant mice. We used the NEX-Cre driver to express BirA in glutamatergic neurons of 6-week male T158M, R106W and control TAVI mice and sorted for exact numbers of excitatory and inhibitory cortical nuclei via FACS. Because we consistently obtained ~1ng of total RNA per 1000 neuronal nuclei, we extracted nuclear RNA (PolyA+ and PolyA-) from 120,000 - 250,000 nuclei per sample to meet the minimum input requirements for Illumina TruSeq Total RNA library preparation with RiboZero depletion. Stranded nuclear RNA-seq (nucRNA-seq) libraries were sequenced 100PE to a mean depth of 30 million read-pairs per replicate, and 90-95% of total reads could be uniquely mapped to the mouse Ensembl GRCm38/mm10 genome assembly due to substantially lower levels of ribosomal RNAs in the nucleus relative to the cytoplasm. We sequenced four biological replicates per genotype and cell type, and replicates for each condition were well correlated (mean r = 0.975) (Figure 3.3E). We

found that the majority of nucRNA-seq reads mapped to the gene body, with approximately 73.9% mapping to introns, 15.6% mapping to exons and the remaining 10.5% mapping to intergenic regions. In total RNA preparations at steady state, intron-mapped reads originate from chromatin-associated primary transcripts that are in the process of being transcribed by RNA polymerase at the time of isolation and are considered a proxy for transcriptional activity. Given that MeCP2 modulates gene expression in the nucleus, these nucRNA-seq datasets afford a unique opportunity to study the effects of *Mecp2* missense mutations on cell type-specific gene expression profiles at the transcriptional level. Hence all genic-mapped reads (defined as total intron- and exon-mapped reads across the gene body) were used for subsequent transcriptomic analyses.

Nuclear Transcriptomes Reveal Functionally Relevant Transcriptional Dynamics Between Cell Types

We first analyzed gene expression profiles from control excitatory and inhibitory neuronal populations to confirm that a transcriptomic analysis of nuclear RNA using genic-mapped reads was informative. Unsupervised hierarchical clustering of mapped read counts revealed that replicate nuclear transcriptomes were strongly correlated and faithfully segregated according to cell type (Figure 3.5A). Genic-mapped reads readily identified highly expressed genes specific to individual cell types, many of which are obscured in traditional mRNA-seq datasets derived from cellularly heterogeneous cortical tissue (Figure 3.3G). Using an FDR < 0.05 as the sole cutoff, we identified 9,379 differentially expressed genes (DEGs) between the two neuronal populations (Table 3.1 and Figure 3.3F), the majority (86.9%) of which comprised protein-coding genes. Approximately half of these genes cell type-enriched genes were associated with Gene Ontology (GO) terms expected of glutamatergic pyramidal cell function, whereas the remaining genes were highly enriched for GO terms consistent with the role of metabolically active cortical interneurons (Figures 3.6A-B). Surprisingly, more than 90% of cell type DEGs exhibited < 3 $\log_2$ fold change and were more likely to be actively expressed in both cell types, albeit to varying degrees. This

suggested that functional differences between excitatory and inhibitory neuronal populations of 6-week old mice may be influenced by small differences in the expression of genes, and we questioned whether this was reflected among GO annotations. We thus sorted significant GO categories by the distribution of genes with low (FC < 1), moderate (1 ≤ FC < 3) or high (FC ≥ 3) $\log_2$ fold enrichment and observed that certain distributions were indeed associated with particular cellular functions (Figures 3.6A-B). Genes that are highly enriched in excitatory neurons (FC ≥ 3), including the neurotrophins *Bdnf* and *Gdnf*, are associated with functions related to secreted extracellular and transmembrane proteins involved in axonal projection and neuronal development. In contrast, genes that are lowly enriched to excitatory neurons (FC < 1) are more associated with intracellular kinase signaling cascades (*Camk4a*) and cytoskeletal remodeling. Genes related to cell adhesion, calcium signaling and other synapse-related functions (*Nrxn1*, *Npas4*) exhibited no strong bias across varying distributions of fold change. Within inhibitory neurons, genes that are lowly enriched (FC < 1) were strongly associated with energy regulation and cellular metabolism in mitochondria. However, highly enriched genes in inhibitory neurons (FC ≥ 3) were also associated with secreted extracellular proteins such as *Ngf* and *Reln*.

Recent studies have demonstrated that in contrast to other genes, neuronal genes tend to be exceptionally long, rendering them susceptible to transcriptional deficits associated with neurodevelopmental and neuropsychiatric disorders, including RTT (Gabel et al., 2015; Sugino et al., 2014; Zylka et al., 2015). Interestingly, we noticed that the excitatory-enriched genes (EXC median = 50.6kb), as well as their respective GO categories, had significantly longer median gene lengths than their inhibitory equivalents (INH median = 21.2kb) (Figures 3.6A-C). This suggests that different neuronal cell types may emphasize the transcription of different gene lengths as a consequence of cellular function. To elucidate how neuronal gene lengths behave with respect to excitatory and inhibitory neuronal gene transcription, we measured the linear correlation between gene lengths and detectable FPKM measurements for all Ensembl-annotated protein-coding genes. Linear and local regression analysis revealed that longer gene lengths were positively associated with increasing gene expression in both cell types, as expected of the

neuronal lineage (Figures 3.6F-G). Pearson correlations among biological replicates were significantly weaker within inhibitory neurons when compared to excitatory neurons, especially among the most active genes that likely comprise the bulk of the cellular transcriptome. When we partitioned the actively expressed genes by cell type enriched genes, we found an anti-correlation between gene length and gene expression among inhibitory-enriched genes in both cell types (Figures 3.6H-I). However, among excitatory-enriched genes, as well as constitutive genes that were not differentially expressed across both cell types, excitatory neurons transcribed increasingly longer genes at higher expression levels, and this correlation was absent in inhibitory neurons. Given that actively expressed genes largely overlap between excitatory and inhibitory neurons and hence display a similar cumulative distribution of gene lengths (Figures 3.6D-E), the cell type-dependent correlations among cell type-enriched gene lengths and expression levels are unlikely to be artifacts associated with the FPKM normalization of read counts. We thus conclude that neuronal cell types display different transcriptional dynamics with regards to gene length in order to carry out cell-specific functions. Therefore, the combined use of intron- and exon-mapped reads from nuclear RNA can yield unique insights about the transcriptional status of neurons *in vivo*.

Protein-Coding Genes are More Severely Affected in R106W than T158M Mice

We next compared nuclear gene expression profiles between age-matched mutant (T158M, R106W) and control (TAVI) mice to identify and characterize differentially expressed genes associated with the appearance of RTT-like phenotypes (Figure 3.7A, summarized in Table 3.1). From both excitatory and inhibitory neurons, we identified a combined total of 223 T158M and 1,069 R106W unique DEGs in 6-week old mutant mice, indicating that the number of misregulated genes positively scales with the severity of the *Mecp2* mutation. *Mecp2* was not among the list of differentially expressed genes in either cell type, confirming that the missense mutations do not affect *Mecp2* expression at the transcriptional level. Between the two cell types, we identified more DEGs in the inhibitory neurons of R106W mice and in the excitatory neurons

of T158M mice (Figure 3.7B). Although the predominantly affected cell type is disparate between *Mecp2* mutant lines, the number and percentage of T158M DEGs that encompass R106W DEGs is greater in inhibitory (n_{shared} = 109 genes, 74.7% of T158M DEGs; Hypergeometric p = $1.98e^{-156}$) than excitatory (n_{shared} = 75 genes, 38.1% of T158M DEGs; Hypergeometric p = $3.34e^{-96}$) neurons. Notably, all of the genes shared between both genotypes were misregulated in the same direction and likely reflect more direct targets of MeCP2 (Figure 3.7C). The Ensembl genome annotation revealed that ≥90% of MeCP2 DEGs comprise protein-coding genes, which was significantly higher than the percentage of protein-coding genes from genomic (60.4%), actively expressed (77.7-78.3%) and cell type-enriched (86.2-87.7%) gene distributions (Figure 3.5C). Given that protein-coding genes are overrepresented among transcriptional deficits associated with mutant MeCP2, we excluded non-coding genes from further analysis.

The most common feature associated with MeCP2-dependent transcriptional dysregulation is the subtle changes in gene expression, and this was reflected among excitatory and inhibitory neurons of both *Mecp2* mutant mice (Table 3.1). However, we questioned whether MeCP2 R106W impacts gene expression to a greater extent than MeCP2 T158M, and we tested this by comparing the fold change distribution amongst differentially expressed genes between *Mecp2* mutant mice. To account for the disproportionate number of DEGs, we limited our analysis to protein-coding genes that were shared across genotypes. Among this subset, the median fold change was consistently and significantly higher in R106W mice, and upregulated and downregulated genes were similarly affected by the severity of the *Mecp2* mutation in both cell types (Figure 3D). The differences between both mutant genotypes were more striking in inhibitory neurons, and a cumulative distribution of T158M and R106W fold change values confirmed a significant rightward shift for the R106W curve across all shared genes in inhibitory neurons (Kolmogorov Smirnov p = 1.316e-5) (Figure 3.7E). R106W values in excitatory neurons also displayed a rightward shift relative to T158M, but the difference was not as robust (Kolmogorov Smirnov p = 0.105). Together, these data suggest that the number of differentially

expressed genes and the degree to which genes are misregulated is influenced by *Mecp2* mutations that differentially impact MeCP2 affinity for methylated DNA.

MeCP2 DEGs Are Typically Low Expressing Cell Type-Enriched Genes

We next examined MeCP2-dependent gene expression changes within the context of how genes are typically expressed in excitatory and inhibitory neurons. Most genes were selectively misregulated in a single cell type as only 20 genes (6.2% of total T158M) and 114 genes (10.7% of total R106W) were misregulated in both excitatory and inhibitory neurons (Figure 3.7F). This degree of overlap is relatively low when compared to the 8,197 protein-coding genes (74.2% of total) that are actively expressed in both cell types. Because the transcriptional changes associated with mutant MeCP2 were cell-specific, we questioned whether there was an association between MeCP2 DEGs and cell type-enriched genes. In each instance, the majority of MeCP2 DEGs (approx. 70-80%) were significantly associated with cell type-enriched genes (Chi-square $p < 2.2e-16$ per comparison; Table 3.1 and Figure 3.7G). In excitatory neurons, ~50% of all T158M and R106W DEGs were classified as excitatory-enriched genes. In inhibitory neurons, T158M DEGs exhibited a stronger enrichment for excitatory-enriched genes, whereas R106W DEGs were equally enriched for both excitatory- and inhibitory-enriched genes (Figure 3.7G).

Recent studies have linked MeCP2 function to gene repression that is coincident with gene length across multiple *Mecp2* mouse models (Gabel et al., 2015). Given that we established an association between cell type-enriched genes and gene lengths in excitatory and inhibitory neurons (Figure 3.6E-F), we next assessed the lengths of differentially expressed protein-coding genes in T158M and R106W mice. MeCP2 DEGs were cumulatively longer than active expressed genes within their respective cell type (Figure 3.8A-B). Among the four lists of MeCP2 DEGs, three had median gene lengths >80kb, with 43-51% of DEGs having gene lengths exceeding 100kb. However, R106W DEGs in inhibitory neurons were significantly shorter with a median length of 40.5kb, and only 26% of genes had gene lengths greater than 100kb. We

partitioned DEG lengths by cell type enriched genes and found that the constitutive and inhibitory-enriched genes, which are significantly shorter than excitatory-enriched genes, were driving the reduced lengths of R106W DEGs in inhibitory neurons (Figure 3.8B).

Curiously, we did not detect differences in gene lengths among constitutive, excitatory- and inhibitory-enriched genes with the remaining MeCP2 DEGs despite clear differences among active expressed genes. Given that cell type-enriched genes are associated with different gene lengths that vary by level of expression in each cell type (see Figure 3.6E-F), we hypothesized that MeCP2 DEGs reflected genes that are typically expressed at low levels where gene lengths are more likely to be equal between constitutive, excitatory- and inhibitory-enriched genes. We thus assessed the relative expression levels of these genes by comparing the $\log_2$ FPKM distribution of MeCP2 DEGs in control TAVI neurons to the distribution of active expressed genes within each cell type (Figure 3.8C). The FPKM of most T158M and R106W DEGs fell below the median of actively expressed genes in both cell types, with many DEGs falling below our cutoff for active expressed genes altogether. This suggests that MeCP2 DEGs are genes that are either lowly expressed or infrequently transcribed, regardless of the severity of the mutation. Interestingly, active expressed genes comprised a higher percentage of R106W DEGs (EXC = 69.2%, INH = 43%) than T158M DEGs (EXC = 52.5%, INH = 40.2%) and were primarily downregulated in R106W mice in both excitatory and inhibitory neurons (Wilcoxon Rank Sum two-tailed p-value, EXC = 8.878e-09, INH = 7.208e-07). In contrast, we did not detect a significant difference in the expression level of upregulated genes between genotypes in either cell type (Wilcoxon Rank Sum two-tailed p-value, EXC = 0.994, INH = 0.9976).

We further tested the association between MeCP2 DEGs and expression levels by dividing active expressed genes into four equally sized bins and determined DEG enrichment (Figure 3.8C). DEGs in both mutant genotypes were significantly enriched in the bottom 25th percentile of active expressed genes (Fisher Exact one-tailed p-value (Odds Ratio), T158M EXC = 1.11e-07 (3.1), T158M INH = 2.03e-04 (3.2), R106W EXC = 4.04e-08 (2.0), R106W INH = 1.50e-02 (1.3)). This association was noticeably stronger in T158M neurons in contrast to R106W

neurons, where DEGs were more evenly distributed across multiple expression bins in both cell types. We also observed a negative correlation between DEG enrichment in the 25th expression percentile and the total number of misregulated genes found in each genotype and cell type (Pearson r = -0.982, p = 0.01849). Thus the T158M mutation appears to affect low-expressing genes in contrast to the R106W mutation, which affects genes found at higher expression levels in addition to low-expressing genes. Given the severity of the phenotypes with respect to the T158M and R106W mutations, this suggests that active expressed genes become more frequently misregulated as MeCP2 function becomes increasingly compromised and the number of DEGs increases in both excitatory and inhibitory neurons.

Upregulated and Downregulated Genes Are Associated With Discrete Gene Functions

Because downregulated genes accounted for the differences in FPKM expression level between T158M and R106W DEGs, we questioned whether upregulated or downregulated DEGs were associated with functionally distinct groups of genes and, if so, whether these functions would be unique to each cell type. This was supported by the observation that the direction in which MeCP2 DEGs were predominantly misregulated was coincident with cell type. Excitatory neurons encompassed more downregulated genes whereas inhibitory neurons encompassed more upregulated genes, and this trend was preserved among shared DEGs between both mutant genotypes (n_{EXC} = 69, 65.2% downregulated; n_{INH} = 107, 64.5% upregulated). Furthermore, upregulated and downregulated genes exhibited distinct preferences for excitatory- or inhibitory-enriched genes in both R106W (Fisher's Exact Test, EXC two-tailed p = 6.057e-13, INH two-tailed p = 3.09e-04) and T158M mice (Fisher's Exact Test, EXC two-tailed p = 3.479e-02; Figure 3.8D).

Among excitatory neurons, we found that excitatory-enriched genes were preferentially downregulated and inhibitory-enriched genes were preferentially upregulated in both mutant genotypes (Figure 3.8D). We further characterized these preferences by performing Gene Set Enrichment Analysis (GSEA) using an FDR cutoff < 0.1 to identify gene functions associated with

upregulated and downregulated genes. GSEA revealed that the majority of upregulated genes in both mutant genotypes were primarily associated with transcriptional regulation (Figure 3.7H-I). Examples of the leading edge genes within these functional annotations included DNA-binding transcriptional activators (*Fos*, *Irf8*, *Egr1*, *Ikzf2*, *Barx2*), co-activators (*Sertad1*), and repressors (*Tcf7l1*, *Tshz1*, *Bhlhe40*, *Btg2*), as well as chromatin remodelers (*Eya1*, *Eya4*, *Ezh2*) and membrane associated receptors that participate in intracellular signaling cascades with transcriptional outcomes (*MET*, *Bmpr1b*, *Htr1b*). In contrast, gene functions associated with downregulated genes were exclusively found within R106W excitatory neurons and were highly enriched for genes encoding post-synaptic membrane proteins, including Na^{2+} (*Scn3b*, *Scn7a*,), K^{+} (*Kcna6*, *Kcnip2*, *Kcnip4*, *Kcna1b*, *Kcnn3*, *Kcnk13*) Ca^{2+} (*Cacng2*, *Cacna1g*, *Cacng4*), and Cl^{-} channels (*Gabra3*, *Gabra5*, *Glra3*), synaptic scaffolding proteins (*Dlgap2*, *Dlgap3*, *Dlgap4*, *Shank2*, *Nrxn2*), and ionotropic glutamate recepeptors (Grin3a, Grik5), among others (Figure 3.7H).

Surprisingly, R106W inhibitory neurons demonstrated the same underlying transcriptional preference for upregulated inhibitory-enriched genes and downregulated excitatory-enriched genes (Figure 3.8D). A number of chromatin remodelers and activity-dependent transcriptional activators were detected, including those that were upregulated (*Creb5*, *Etv1*, *Tead1*, *Eya1*) and downregulated (*Nr4a3*, *Npas1*, *Arc*, *Ube2b*). However, GSEA functional annotations of MeCP2 DEGs were notably distinct from excitatory neurons due to the different transcriptional environment of interneurons. Gene functions associated with upregulated genes were related to cellular metabolism and respiration, DNA damage and repair (*Mgmt*, *Nabp1*, *Parpbp*, *Uvssa*), cell surface membrane and secreted proteins (*Reck*, *Slit1*, *Sema5a*, *Fzd6*, *Shh*), cytoskeletal/scaffold proteins (*Rhoj*, *Palld*), and proteins that mediate intracellular signaling cascades (*Nsmf*, *Gfra1*, *Abca1*). Although downregulated gene functions were not identified using our established FDR cutoff, a more permissive cutoff (FDR < 0.25) revealed gene functions associated with post-synaptic membrane functions, such as *Shank1*, *Reln*, *Homer1*, and *Frmpd4*.

Transcriptional Dysregulation in T158M Mice Increases With Age

RTT is a progressive disorder with clinical symptoms and phenotypes that worsen over time. Although differentially expressed genes are known to increase in *Mecp2*-null mice with age, we questioned whether the same would be observed for a hypomorphic mutation such as T158M (Jordan et al., 2007). We thus aged T158M an additional 6 weeks and transcriptionally profiled excitatory and inhibitory neuronal nuclei from 12 week old mice. Although we observed an increase in the number of differentially expressed genes with age, the magnitude of the increase was surprising. T158M DEGs increased from 197 to 5889 coding and non-coding genes in excitatory neurons, and from 758 to 8086 coding and non-coding genes from inhibitory neurons (Figure 3.9A).

In addition to the increased number of DEGs, aged T158M mice exhibited extensive overlap in DEGs shared across cell types, in contrast to younger T158M and R106W animals (Figure 3.9B). The distribution of cell type-enriched gene lengths among aged T158M DEGs closely matched the distribution across active expressed genes, which suggests that the majority of T158M DEGs were affecting genes expressed at higher levels. We found that aged T158M DEGs misregulated across both cell types were typically expressed at higher levels than genes misregulated in a single cell type, which were expressed at lower levels (Figure 3.11C). Another striking feature of aged T158M DEGs was the strong association that was observed between the direction of transcriptional change and gene length. Many long genes > 100kb in length were downregulated, whereas many genes ≤ 100kb in length were upregulated (Figures 3.9B-C), which conflicts with recent studies implicating MeCP2 in length-dependent repression of long genes and is further addressed in the proceeding section.

Finally, we compared differentially expressed genes between 6 week T158M and R106W neurons and 12 week T158M neurons. In both excitatory and inhibitory neurons, the majority of DEGs at 6 weeks were also detected among DEGs at 12 weeks for each cell type, and the degree of change was consistently larger in aged T158M excitatory and inhibitory neurons relative to 6 week neurons (Figures 3.9D-F). Interestingly, aged T158M DEGs overlapped with a

larger proportion of DEGs in R106W mutant neurons at 6 weeks (EXC: 57.3%, INH: 50.7%) than T158M mutant neurons at 6 weeks (EXC: 44.4%, INH: 23.5%). This indicates that *Mecp2*-mutant neurons are more likely to share DEGs based on the degree of phenotypic severity, rather than simply being of the same genotype or mutation.

Global Transcription of Long Genes is Reduced and Leads To Post-transcriptional Upregulation of Long Gene RNA

Several recent studies have demonstrated that MeCP2 is involved in the transcriptional repression of long genes, which are preferentially upregulated in the neurons of multiple MeCP2-deficient model systems (Gabel et al., 2015; Sugino et al., 2014). We sought to reproduce this result in our own mutant mice, focusing specifically on the R106W mutation. As a positive control, we isolated whole cell RNA from 6-week cortical tissue as described (Gabel et al., 2015), as well as unsorted nuclear RNA from the same biological samples. After sequencing and mapping reads, we sorted and binned protein-coding genes according to gene length and measured the mean fold change between our control and mutant mice for both whole cell and nuclear RNA fractions. We observed a similar length-dependent increase in the mean expression of long genes from whole cell RNA as previously described (Gabel et al., 2015). However, we discovered a striking inversion of this trend with nuclear RNA, which revealed that genes ≤ 100kb in length were upregulated and genes > 100kb in length were downregulated in a length-dependent manner (Figure 3.10A). To confirm that the gene expression changes observed in nuclear RNA fractions represented true transcriptional changes, we used GRO-seq to profile and map nascent RNA transcripts from isolated cortical nuclei of 6-week TAVI and R106W mice, and repeated the length-dependent analysis (Core et al., 2008). Within mutant nuclei, we found a predominant increase in nascent transcription for genes ≤ 100kb in length, and a predominant length-dependent decrease in nascent transcription of genes > 100kb in length (Figure 3.10B), similar to what was observed in nuclear RNA. Genes that were upregulated in nascent and nuclear RNA fractions were cumulatively shorter relative to genes upregulated in whole cell RNA (Kolmogorov-

Smirnov two-tailed p-value < 2.2e-16 for both comparisons), and the inverse was observed for downregulated genes (Kolmogorov-Smirnov two-tailed p-value < 2.2e-16 for both comparisons, Figure 3.10C). We observed a weak but positive correlation among $\log_2$ fold changes for each gene across both nascent and whole cell RNA fractions (Pearson's r = 0.134) that indicates that changes in expression at the transcriptional level are minimized at the post-transcriptional level (Figure 3.10D). Hence we find that global gene expression changes in *Mecp2*-mutant mice differ between subcellular compartments and that nuclear RNA more accurately reflects the transcriptional landscape associated with MeCP2 deficiency.

Gene Length-Dependent Transcriptional Differences Correlate with Disease Severity

We next asked if we could identify global length-dependent transcriptional changes that correlate with phenotypic severity pertaining to age and genotype, analogous to what was found for FDR-controlled MeCP2 DEGs (see Figures 3.6 and 3.7). Similar to nascent and nuclear RNA fractions from whole cortical tissue, sorted nuclear RNA from both excitatory and inhibitory mutant neurons displayed length-dependent upregulation of genes ≤ 100kb in length and downregulation of genes > 100kb in length. Although we observed a significant increase in the overall degree of gene misregulation from 6 to 12 weeks of age, we did not detect a notable difference in length-dependent transcriptional changes between T158M and R106W neurons at 6 weeks of age in either cell type (Figures 3.11G and 3.11I).

We reasoned that differences in Pol II transcriptional activity at the gene body might underlie transcriptional differences between T158M and R106W mutant neurons at the same biological age, which would require an ability to detect nascent RNA transcripts that correlate with transcriptional activity. We thus used our GRO-seq dataset to identify genes in sorted neurons that met the aforementioned criteria (Figure 3.10E) and repeated the length-dependent analysis of 6-week and 12-week samples using this gene filter. Between 6-week T158M and R106W excitatory and inhibitory neurons, GRO-seq-filtered genes revealed a striking difference in mean fold change among downregulated genes > 100kb in gene length (Figures 3.10F-I). The

difference in magnitude originated from T158M mutant neurons, whose fold changes were considerably smaller than R106W mutant neurons after filtering (Figures 3.10G, 3.10I, 3.11H, and 3.11J). Furthermore, this effect was only observed among samples at the same biological age and did not have a noticeable impact on fold changes from 12-week T158M samples. Hence, transcriptional downregulation of genes >100kb in gene length specifically correlates with differences in the severity of *Mecp2* missense mutations at the methyl-CpG binding domain.

A Female RTT Mouse Model Reveals Cell and Non-Cell Autonomous DEGs

Among heterozygous female mice bearing a mutant *Mecp2* allele, random X-inactivation leads to both genetic mosaicism among neuronal populations and phenotypic variability. As a result, the majority of RTT molecular research is performed using hemizygous male mice, whereas the majority of studies involving female mice are limited to behavioral assessments and immunohistological correlates (Lyst and Bird, 2015). The ability to label and isolate WT and mutant neurons from heterozygous female mice for transcriptional profiling may yield new molecular insights into RTT etiology and MeCP2 function in the CNS.

As a proof-of-principle, we generated heterozygous TAVI ($Mecp2^{Tavi/+}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$), T158M ($Mecp2^{Tavi/T158M\text{-}Tavi}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$), and R106W ($Mecp2^{Tavi/R106W\text{-}Tavi}$;$R26^{cBirA/+}$;$NEX^{Cre/+}$) female mice that each carried one copy of the *Mecp2-Tavi* allele. We aged female mice to 18 weeks for phenotypic characterization prior to FACS isolation of cortical nuclei. Similar to male mice, the phenotypic scores of TAVI female mice remained comparatively low at 3 and 4 months of age, whereas T158M and R106W female mice both displayed significantly higher phenotypic scores (Figure 3.12B). This confirms that the *Mecp2-Tavi* allele is functionally equivalent to the wild-type *Mecp2* allele and that an increase in phenotypic score only occurs when bearing one the *Mecp2* missense mutations.

Using FACS isolation of cortical nuclei, we could distinguish wild-type (denoted as $TAVI_{WT}$, $T158M_{WT}$, $R106W_{WT}$) and mutant (denoted as $T158M_{MUT}$, $R106W_{MUT}$) populations derived from a single female cortex because of the reduced levels of MeCP2 protein within

mutant neurons (Figures 3.12A and 3.13A). Interestingly, we found that the amount of biotinylated MeCP2-Tavi detected in mutant neurons was significantly higher in T158M than R106W mice (29.1% vs. 20.8%, Student's t-test; p = 0.038; Figure 3.12C). This indicated that MeCP2 T158M is present at higher protein expression levels than MeCP2 R106W in the neurons of older female mice and is reminiscent of mutant MeCP2 protein expression levels in younger male mice. Among excitatory neurons, TAVI, T158M, and R106W mice displayed similar mean X-inactivation ratios (53.1% $TAVI_{WT}$, 42.6% $T158M_{WT}$, 49.2% $R106W_{WT}$) and did not display skewed X-inactivation > 75% (Figure 3.12D).

Nuclear RNA-seq from these sorted populations revealed striking differences between female mutant mice. We first identified differentially expressed genes in $T158M_{WT/MUT}$ and $R106W_{WT/MUT}$ excitatory neuronal populations by comparing them to $TAVI_{WT}$, revealing a total of 526 and 678 unique protein-coding genes, respectively (Table 3.1). By comparing the distribution of DEGs detected between WT and mutant neurons, we found that 5.3% of T158M DEGs and 43.4% of R106W DEGs were non-cell autonomous gene expression changes, revealing a surprising mutation-specific transcriptional vulnerability of WT neurons in RTT (Figure 3.12E and 3.13B). We further visualized the differences between sorted female populations using principal component space to plot the first two major axes of variation (PC1 vs. PC2) (Figure 3.12F). PC1 sufficiently separated R106W-derivied populations from the remaining genotypes and likely reflected the presence of non-cell autonomous DEGs among both wild-type and mutant neurons. PC2 further separated neuronal populations by *Mecp2* allele status (wild-type vs. mutant neurons) irrespective of genotype, suggesting that *Mecp2* mutants induce cell autonomous transcriptional changes that distinguish mutant neurons from neighboring wild-type neurons. Against PC1 and PC2, $T158M_{WT}$ neurons closely resembled that of $TAVI_{WT}$, which indicates that adjacent positioning to mutant neurons is not sufficient to generate the degree of non-cell autonomous transcriptional changes apparent in R106W mice. As PC1 accounts for twice the amount of variation as PC2 and similarly affects both wild-type and mutant neurons, these data

suggest that non-cell autonomous transcriptional changes in R106W female mice may be extensive.

Given that $R106W_{MUT}$ clusters away from $R106W_{WT}$, we questioned whether non-cell autonomous DEGs represented a molecularly distinct group from cell autonomous DEGs. We further assessed the distribution of fold changes among cell and non-cell autonomous DEGs in R106W mice and observed that cell autonomous DEGs, which are more likely mediated by mutant MeCP2, had significantly smaller fold changes than non-cell autonomous DEGs (Wilxoxon Rank Sum, p = 1.077e-07) (Figure 3.13C). However, we did not detect a difference in absolute fold change between non-cell autonomous DEGs that are present in both $R106W_{WT}$ and $R106W_{MUT}$ neurons (Wilcoxon Signed Rank; p = 0.2461). Upregulated cell autonomous DEGs were also cumulatively longer in gene length than non-cell autonomous DEGs (Figure 3.13D). GSEA functional enrichment revealed that non-cell autonomous gene expression changes were enriched for several early and late response genes that are known to link neuronal activity to gene expression and affect synaptic plasticity, including *Bdnf* and *Igf1*, which are known to stimulate signaling pathways that lead to the activation of immediate early gene expression such as Npas4 and Fos (Figure 3.13E). The majority of molecules that were upregulated appeared to be neuroprotective; thus these non-cell autonomous changes may be promoting cell health, likely in response to abnormal neuronal network activity. In contrast, cell autonomous DEGs were enriched for upregulated transcriptional regulators and downregulated G-protein coupled receptors, similar to what we observed in 6 week male excitatory neurons (Figure 3.12J). Thus non-cell autonomous and cell autonomous DEGs reflect functionally different cellular processes.

We identified 194 protein-coding genes shared between T158M and R106W female mice, 193 of which were misregulated in the same direction (Figure 3.12G). Among 194 genes, 149 (76.8%) were considered cell autonomous in both mutant mice, whereas 9 genes (4.6%) were considered non-cell autonomous in both mutant mice, indicating that cell autonomous transcriptional changes are more likely to shared across genotypes (Fisher's Exact Test, p = 2.43e-05, Figure 3.12H). The remaining 36 genes (18.5%) were found to be cell autonomous in

one genotype and non-cell autonomous in the other. The majority of these genes (n = 34) were cell autonomous in T158M mice and non-cell autonomous in R106W mice and thus reflect a limitation of detecting non-cell autonomous DEGs in the absence of transcriptionally vulnerable $T158M_{WT}$ neurons.

We next sought to determine whether shared DEGs were more severely affected in R106W than T158M female mice, similar to what was observed with 6-week old male mice. Among shared DEGs in 18-week female mice (n = 185 genes, excluding genes that are exclusively non-cell autonomous), protein-coding genes were indeed more severely affected in R106W than T158M mice (Wilcoxon Signed Rank; p = 0.02711, Figure 3.12I). In contrast to younger male mice, this effect was limited to upregulated genes (Wilcoxon Signed Rank; p = 0.002995), whereas downregulated genes were similarly affected across genotypes (Wilcoxon Signed Rank; p = 0.3952). We next tested whether these genotype-specific differences occur in a cell autonomous fashion, and thus are likely to be mediated by MeCP2. Using genes that are considered cell autonomous in both genotypes (n = 149 genes), we did not detect a difference in absolute fold change between T158M and R106W mice (Wilcoxon Signed Rank; p = 0.5272, Figure 3.12I). However, among the 36 genes that are non-cell autonomous in at least one genotype, we again observed that these shared genes were more severely affected in R106W than T158M mice (Wilcoxon Signed Rank; p = 4.593e-05, Figure 3.12I), including both upregulated (Wilcoxon Signed Rank; p = 0.007751) and downregulated genes (Wilcoxon Signed Rank; p = 0.001356).

Finally, GSEA functional analysis of shared genes between T158M and R106W mutant neurons revealed upregulated genes to be associated with transcriptional regulators and downregulated genes to be associated with transmembrane ion channels, strikingly similar to what was observed in younger 6-week male excitatory neurons. Thus, similar pathogenic processes may be occurring in the excitatory neurons of both RTT male and female mice.

DISCUSSION

In this study, we performed a comprehensive transcriptome analysis of FACS-isolated *Mecp2*-mutant neuronal nuclei across different cellular and subcellular contexts. This was accomplished through the development of a novel genetic strategy that allows for cell type-specific biotinylation of Tavi-tagged MeCP2 and RTT-associated mutant variants *in vivo*. Although similar methods have been developed to biotinylate tagged proteins in mice and other eukaryotic systems, biotinylation typically occurs either ubiquitously throughout the organism or is restricted to a single cell type (Boer et al., 2003; Driegen et al., 2005; Kim et al., 2009; Malik and Henikoff, 2010; Ooi et al., 2010). The coupling of biotinylation with Cre-Lox technology powerfully extends the use and flexibility of the Tavi tag *in vivo* for ChIP, protein-protein co-immunoprecipitations and transcriptional profiling in different cell types based on the availability of existing Cre mouse lines (Gong et al., 2007). Furthermore, this system can be expanded to any protein-coding gene because the small size of the Tavi tag (66 nucleotides) makes it ideal for targeted insertion using CRISPER-Cas9 technology (Yang et al., 2013).

Genome-wide Transcriptional Changes in RTT are Post-Transcriptionally Compensated

In contrast to whole cell RNA, which is enriched for cytoplasmic mRNAs subject to post-transcriptional regulation and degradation, nuclear RNA provides a better correlate of transcriptional dynamics in the cell (Bhatt et al., 2012; Djebali et al., 2012; Werner and Ruthenburg, 2015). Because MeCP2 is a predominantly nuclear protein, nuclear RNA from *Mecp2*-mutant mice may reveal transcriptional effects that are obscured from whole cell or translating mRNA preparations. We demonstrated that, in the absence of MeCP2 function using two independent RTT-associated missense mutations, protein-coding genes ≤ 100kb in length are preferentially upregulated while protein-coding genes > 100kb in length are preferentially downregulated in both unsorted and sorted neuronal nuclei, and the degree of misregulation occurs in a length-dependent manner that increases with age. Using GRO-seq to measure levels of nascent RNA synthesis genome-wide, we confirmed that these changes occur at the

transcriptional level. Sorted nuclei also revealed that these changes occur in both excitatory and inhibitory neurons, and the degree of misregulation was considerably smaller in inhibitory neurons than excitatory neurons at 6 weeks of age. We posit that this may somehow relate to the strong positive correlations between gene length and expression level in excitatory neurons that are notably diminished in inhibitory neurons.

These observations run counter to several recent reports that have identified length-dependent upregulation of long genes > 100kb in different neuronal cell types and brain regions of multiple *Mecp2* models, along with the functional interpretation that MeCP2 is biased towards the repression of long genes (Brown et al., 2015; Gabel et al., 2015; Sugino et al., 2014). However, cortical whole cell RNA preparations using our mutant mice were also able to reproduce the same length-dependent upregulation of long genes > 100kb, similar to these studies. We found that mean expression changes in whole cell RNA negatively correlate with that of nuclear and nascent RNA fractions. In addition, whole cell RNA displayed cumulatively shorter downregulated genes and longer upregulated genes in opposition to changes observed in the nucleus. The apparent discrepancy between subcellular compartments suggests that in the absence of MeCP2 function, subtle length-dependent transcriptional changes in the nuclear transcriptome are biologically relevant and likely subject to post-transcriptional compensation that acts to minimize these changes. Although it remains unclear whether such compensation of transcriptional deficits is truly beneficial or pathological in its own right, this mechanism could underlie the stabilization, and at times subtle improvement, of symptoms that patients experience after the initial onset of developmental stagnation and regression associated with RTT (Neul et al., 2010). Identifying the molecular players that underlie this compensation may thus yield a novel class of interventional therapeutics that can be administered prior to or during the initial phase of RTT and minimize its clinical impact.

Transcriptional Profiling in Female RTT Mouse Models

Importantly, our genetic strategy overcomes a previously insurmountable barrier for research into RTT, and allows for the separation and isolation of genetically mosaic neuronal nuclei from heterozygous female mice for transcriptional profiling. In males, excitatory neurons at 6 weeks of age displayed DEGs whose numbers scaled with the molecular severity of the mutation, whereas neurons at 12 weeks of age displayed thousands of DEGs that were significantly length-dependent. Despite the advanced age of the females used in this study (18 weeks), female mutant neurons displayed an intermediate number of DEGs that scaled with the severity of the mutation, although to a lesser extent than what was observed in males. Transcriptional profiling of excitatory neurons from both T158M and R106W heterozygous mice thus supports the notion that female mutant neurons are on a pathological timeline that is distinct from male mutant neurons, which is likely due to the presence of adjacent wild-type neurons. However, it remains unclear to what extent transcriptional changes may be shaped by the availability of wild-type neurons that provide trophic support. Further experiments assessing multiple females that encompass a wide range of X-inactivation ratios at multiple time points, and perhaps across multiple cell types, will be required to resolve this issue.

A significant advantage of this approach is the ability to distinguish cell autonomous effects from molecularly distinct non-cell autonomous effects at the transcriptional level, which may better inform genotype-phenotype correlations among *Mecp2* mutations. Post-mortem brains from RTT patients exhibit reduced dendritic arborization and defects in spine density and morphology, which are recapitulated in *Mecp2*-null animal models and are characteristic of deficits in synaptic plasticity and neuronal network activity (Armstrong, 2005; Chapleau et al., 2009; Goffin et al., 2012; Wang et al., 2013). By comparing two mutations with different molecular severities, we discovered that both wild-type and mutant neurons were susceptible to non-cell autonomous gene expression changes in R106W, but not T158M, mice. Many of these non-cell autonomous changes were associated with upregulated genes that link neuronal activity to gene expression and help promote synaptic plasticity. Two of these genes (*Bdnf* and *Igf1*) encode neuroprotective peptides that improve RTT synaptic, cellular, and behavioral phenotypes and are

currently being tested in clinical trials (Castro et al., 2014; Chang et al., 2006; Khwaja et al., 2014). The absence of these non-cell autonomous changes in T158M mice could indicate that neurons expressing the R106W allele are more prone to synaptic or circuit abnormalities than neurons expressing the T158M allele *in vivo*. Thus, the severe synaptic plasticity deficits that are commonly observed in *Mecp2*-null mice might not be fully recapitulated across the broad spectrum of RTT-associated mutations that may differentially impact synaptic or circuit abnormalities contributing to RTT clinical severity.

	Experiment	Cell Type	Genotype	# DEGs			Upregulated/	Absolute Median Log_2FC		Gene Length (kb)*		Cell Type Specific Enrichment		
				Total	Coding	Non-coding	Downregulated	Median	IQR	Median	IQR	Both	Excitatory	Inhibitory
6 WEEK (Male)	Actively Expressed Genes	Excitatory Neurons	Control	13877	10926	2951	-	-	-	24.9 kb	9.9 - 61.3	46.5%	28.5%	24.9%
		Inhibitory Neurons	Control	10369	8319	2050	-	-	-	25.4 kb	10.3 - 61.5	46.8%	24.3%	28.8%
	Cell Type-enriched Gene Expression	Excitatory Neurons	Control	4593	3958	635	-	0.75	0.41 - 1.46	50.6 kb	21.5 - 110.1	-	100%	-
		Inhibitory Neurons	Control	4783	4194	589	-	0.73	0.45 - 1.47	21.2 kb	8.8 - 53.0	-	-	100%
	MeCP2-dependent Gene Expression	Excitatory Neurons	T158M	197	177	20	63.8% DOWN	0.45	0.33 - 0.62	88.6 kb	33.5 - 173.1	16.4%	49.2%	34.5%
			R106W	425	386	39	61.7% DOWN	0.44	0.34 - 0.61	82.5 kb	36.7 - 146.0	17.6%	51.3%	31.1%
			Shared	75	69	6	65.2% DOWN	see figure 4		109.8 kb	49.9 - 184.1	14.5%	52.2%	33.3%
		Inhibitory Neurons	T158M	146	143	3	62.9% UP	0.45	0.29 - 0.55	107.8 kb	41.4 - 188.7	20.3%	44.8%	35.0%
			R106W	758	697	61	56.2% UP	0.41	0.33 - 0.56	40.5 kb	19.1 - 102.9	29.4%	35.3%	35.3%
			Shared	109	107	2	64.5% UP	see figure 4		112.8 kb	42.9 - 202.8	18.7%	46.7%	34.6%
12 WEEKS (Male)	Actively Expressed Genes	Excitatory Neurons	Control	12232	9768	2464	-	-	-	25.2 kb	10.1 - 61.3	50.4%	28.0%	21.6%
		Inhibitory Neurons	Control	12031	9596	2435	-	-	-	25.6 kb	10.3 - 62.9	49.8%	23.7%	26.5%
	Cell Type-enriched Gene Expression	Excitatory Neurons	Control	3889	3522	377	-	0.62	0.36 - 1.21	39.7 kb	16.4 - 91.9	-	100%	-
		Inhibitory Neurons	Control	4064	3638	426	-	0.6	0.34 - 1.35	37.3 kb	16.2 - 93.9	-	-	100%
	MeCP2-dependent Gene Expression	Excitatory Neurons	T158M	5889	5324	565	50.8% UP	0.65	0.49 - 0.86	28.9 kb	8.7 - 106.1	44.1%	29.8%	26.1%
		Inhibitory Neurons	T158M	8086	7250	836	51.3% UP	0.61	0.42 - 0.86	29.3 kb	9.7 - 94	46.6%	25.1%	28.5%
18 WEEKS (Female)	MeCP2-dependent Gene Expression	Excitatory Neurons	$T158M_{WT}$	42	28	14	64.3% DOWN	0.52	0.49 - 0.79	54.6 kb	27.2 - 112.6	-	-	-
			$R106W_{WT}$	346	294	52	62.2% UP	0.64	0.47 - 0.93	51.4 kb	21.0 - 140.4	-	-	-
			$T158M_{MUT}$	585	516	69	67.2% DOWN	0.49	0.40 - 0.73	51.2kb	20.4 - 113.2	-	-	-
			$R106W_{MUT}$	634	569	65	52.2% UP	0.54	0.43 - 0.83	65 kb	31.7 - 142.1	-	-	-
			$Shared_{MUT}$	207	185	22	65.9% DOWN	see figure 7		77.5 kb	36.2 - 143.6	-	-	-

Table 3.1: List of active expressed genes, cell type-enriched genes, and *Mecp2* differentially expressed genes derived from sorted nuclear RNA in Chapter 3. Protein-coding genes are considered only.

CHAPTER 3 FIGURES

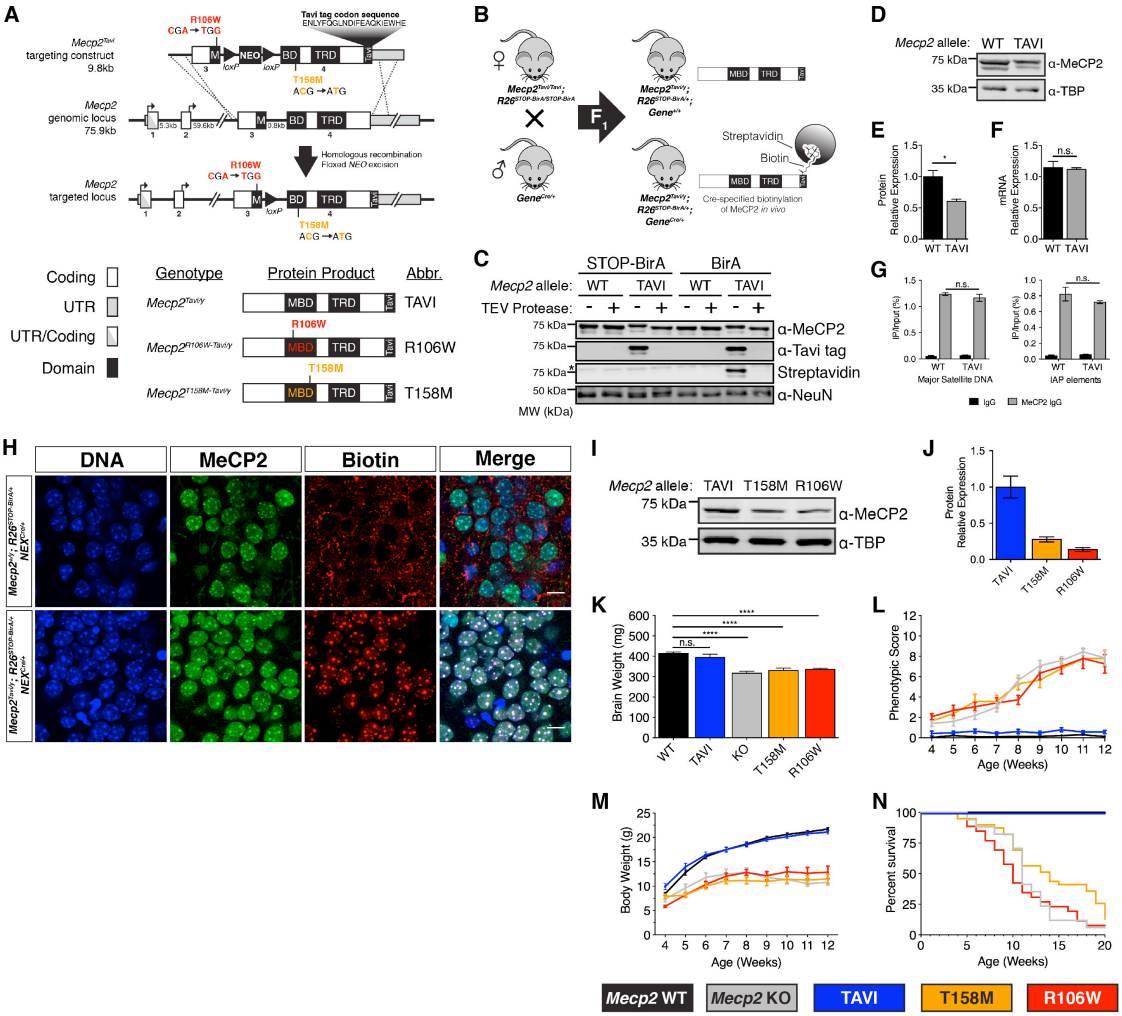

FIGURE 3.1: Design, Utilization, and Characterization of *Mecp2*^{Tavi} mice and Associated RTT Variants

(A) Schematic of the targeting strategy used for generating Tavi-tagged knock-in mice at the endogenous *Mecp2* genomic locus (top). Schematic of translated Tavi-tagged protein products that display the placement of the R106W or T158M missense mutations along with associated genotype and shorthand abbreviations (bottom). MBD, Methyl-CpG Binding Domain; TRD, Transcriptional Repression Domain; NEO, Neomycin cassette.

(B) Diagram of the mouse breeding strategy required to biotinylate the Tavi tag in a Cre-dependent manner.

(C) Representative western blot analysis detailing genotype/conditions under which the Tavi tag is biotinylated using whole brain nuclear extracts treated with or without TEV protease. Blot is

probed with antibodies against the MeCP2 N-terminus, the Tavi tag, and NeuN, and probed with streptavidin for biotin detection.

(D) Western blot comparing MeCP2 protein expression between wild-type untagged and tagged mice at 20 weeks of age. Blot is probed with antibodies against MeCP2 C-terminus and TBP.

(E) Relative protein expression quantification of western blot in **(D)** (n = 3; Student t-test).

(F) Relative expression of MeCP2 mRNA between wild-type untagged and tagged mice at 20 weeks of age using RT-PCR (n = 3; Student t-test)

(G) ChIP-PCR signal of MeCP2 binding at highly methylated repetitive elements using antibodies against the MeCP2 C-terminus or normal rabbit IgG. (n=3, Student t-test).

(H) Representative images depicting immunofluorescent detection of biotinylated MeCP2 and RTT variants in cortical tissue of untagged (WT) and tagged (TAVI, T158M, R106W) male mice at 6 weeks of age using an antibody against the MeCP2 C-terminus and streptavidin for biotin detection. Scale bars represent 10 μm.

(I) Representative western blot comparing MeCP2 relative protein expression levels in the brain between Tavi-tagged control (TAVI) and mutant (T158M, R106W) male mice at 6 weeks of age. Blot is probed with antibodies against the MeCP2 C-terminus and TBP.

(J) Relative protein expression quantification of western blot in **(I)** (n = 4, Student t-test).

(K) Brain weights from untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice at 6 weeks of age (n_{WT} = 20, n_{TAVI} = 11, n_{KO} = 6, n_{T158M} = 6, n_{R106W} = 12;F = 20.05, $P < 0.0001$, one way ANOVA).

(L) RTT-like phenotypic presentation over postnatal age in untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice. Data points over time consist of at least 6 observations each. Total number of mice assessed: n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 15, n_{T158M} = 14, n_{R106W} = 28.

(M) Body weight presentation over postnatal age in untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice. Data points consist of at least 6 observations each. Total number of mice assessed: n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 15, n_{T158M} = 14, n_{R106W} = 28.

(N) Kaplan-Meier survival curve for untagged (WT, KO) and tagged (TAVI, T158M, R106W) male mice (n_{WT} = 31, n_{TAVI} = 23, n_{KO} = 17, n_{T158M} = 39, n_{R106W} = 26; χ^2 = 109.3, df = 4, $P < 0.0001$, Mantel-Cox).

*$P < 0.5$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, n.s. = not significant; All data depicts mean ± SEM. See also **Figure 3.2**.

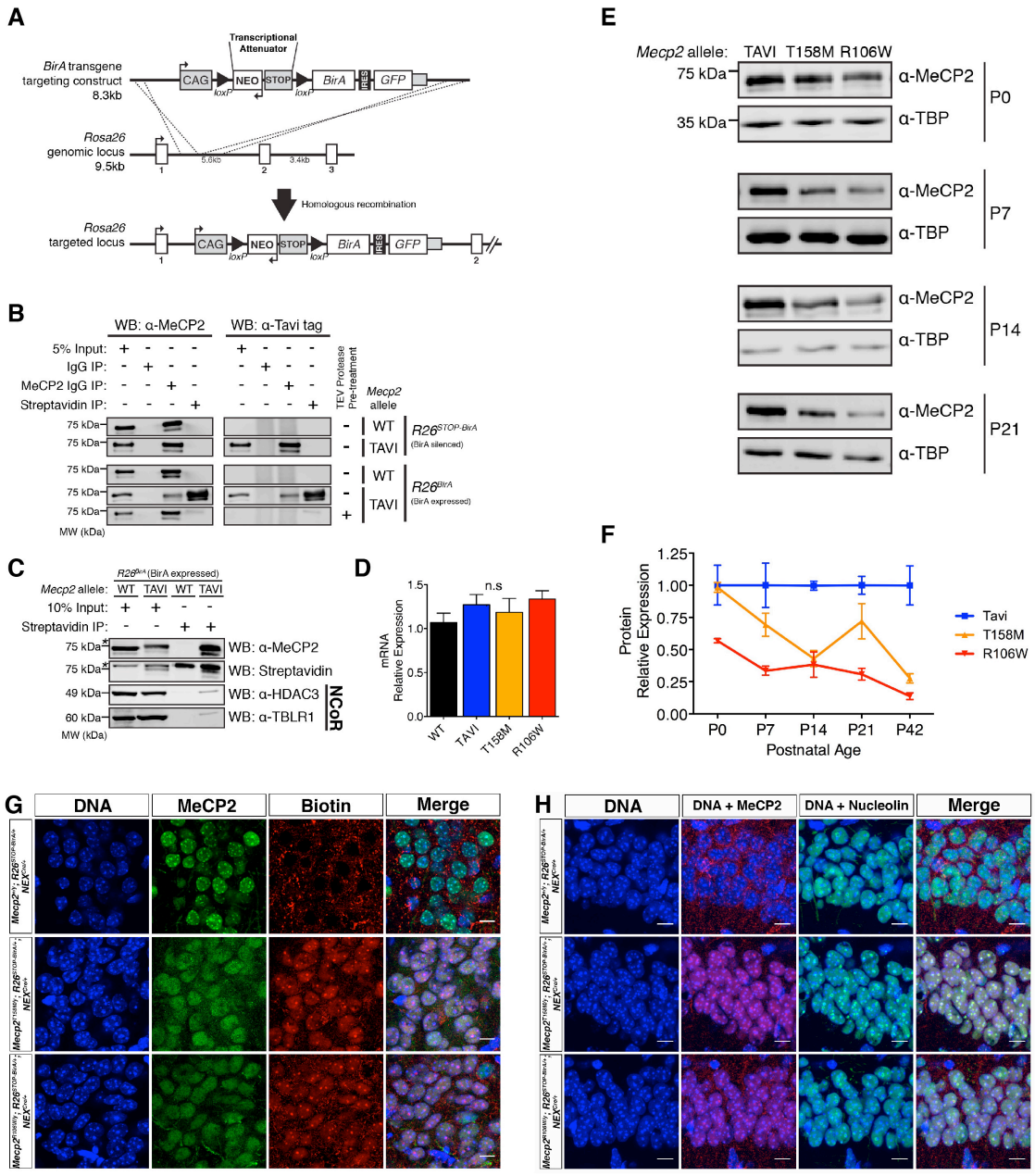

FIGURE 3.2: Extended Data for Figure 3.1

(A) Schematic of the targeting strategy used for generating conditional BirA transgenic mice at the Rosa26 genomic locus.

(B) Western blot detailing genotype/conditions under which the Tavi tag is biotinylated and amenable to be isolated via antibody- or streptavidin-mediated pulldown. Whole brain nuclear extracts treated with or without TEV protease were used. Immunoprecipitation was performed

with antibody against the MeCP2 C-terminus. Blot is probed with antibodies against the MeCP2 N-terminus and the Tavi tag.

(C) Streptavidin-mediated pulldown of biotinylated MeCP2-TAVI and interacting NCoR co-repressor components. Blot was probed with antibodies against the MeCP2 C-terminus, HDAC3, TBLR1, and streptavidin for biotin detection.

(D) Relative expression of MeCP2 mRNA between wild-type untagged and tagged mice at 6 weeks of age using RT-PCR (n=3-4 per genotype, one way ANOVA).

(E) Representative western blot comparing MeCP2 relative protein expression levels in the brain between Tavi-tagged control (TAVI) and mutant (T158M, R106W) male mice at 0, 1, 2, and 3 weeks of age. Blot is probed with antibodies against the MeCP2 C-terminus and TBP.

(F) Quantification of western blots depicted in **(E)** (n = 3-4 per genotype).

(G) Representative images depicting immunofluorescent detection of biotinylated MeCP2 RTT variants in cortical tissue of untagged (WT) and tagged (T158M, R106W) male mice at 6 weeks of age using an antibody against the MeCP2 C-terminus and streptavidin for biotin detection. Scale bars represent 10μm.

(H) Representative images depicting immunofluorescent spatial colocalization of biotinylated MeCP2 RTT variants and nucleoli in cortical tissue of untagged (WT) and tagged (T158M, R106W) male mice at 6 weeks of age using an antibody against nucleolin and streptavidin for biotin detection. Scale bars represent 10μm.

*$P < 0.5$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, n.s. = not significant.

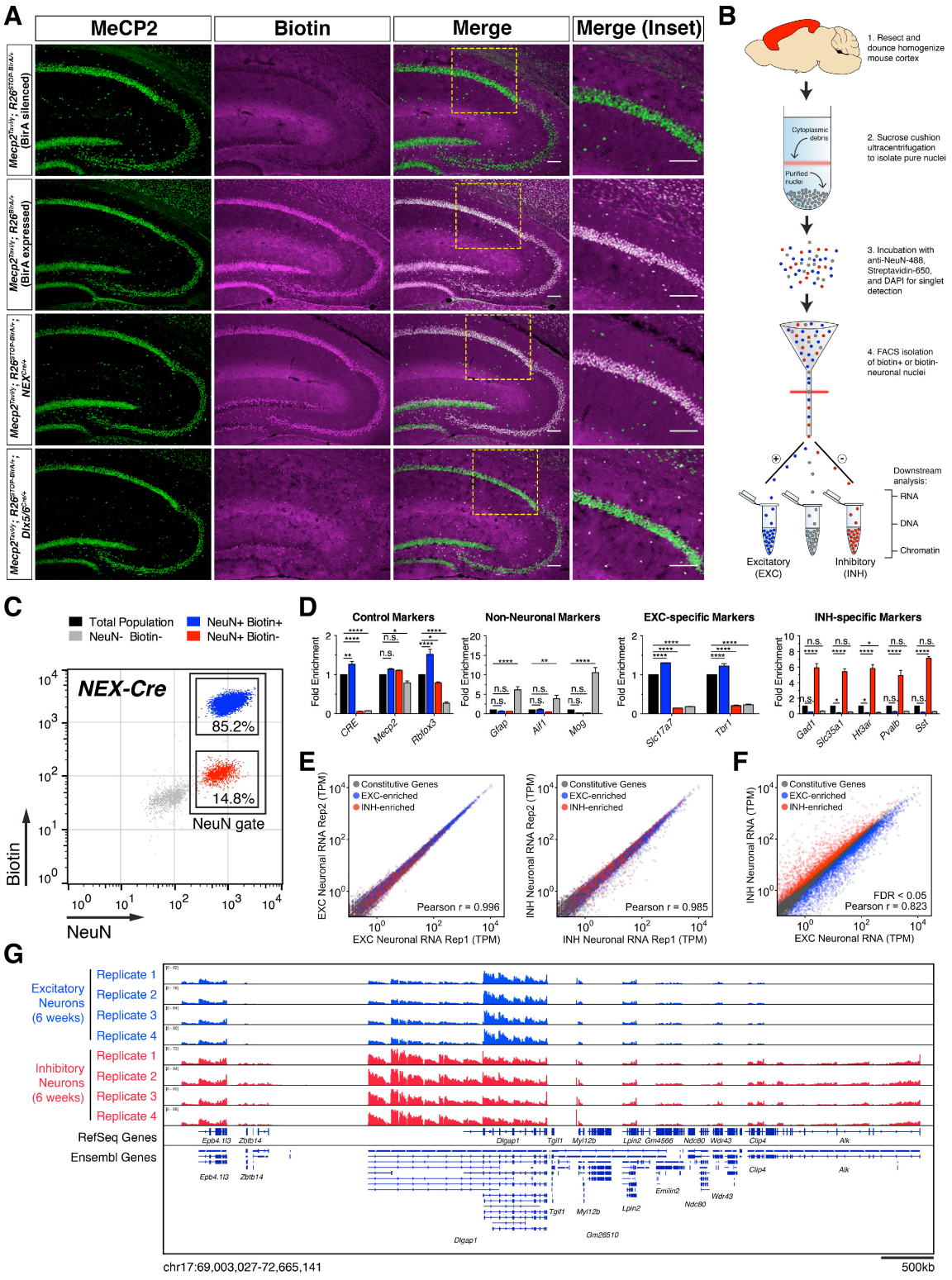

FIGURE 3.3: Cell type-specific transcriptional profiling of neuronal nuclei

(A) Representative immunofluorescent detection of biotinylated MeCP2-Tavi protein within Cre-specified neuronal populations of the mouse hippocampus using an antibody against the MeCP2 C-terminus and streptavidin for biotin detection. See **Figure 3.4A-B** for additional brain regions. Scale bars represent 100μm.

(B) Schematic of cortical nuclei isolation and preparation for Fluorescence-Activated Cell Sorting (FACS).

(C) FACS analysis of labeled cortical nuclei populations. Data shown is representative of nine independent experiments using NEX-Cre mice. Percentages indicate the mean distribution of neurons that are biotin+ (excitatory; 85.2% ± 0.35) or biotin- (inhibitory; 14.8% ± 0.35). See **Figure 3.4I-J** for reverse experiment.

(D) RT-PCR validation of FACS-isolated populations depicted in **(C)**.

(E) Pearson correlation of replicate sequencing libraries using nuclear RNA from FACS-isolated populations depicted in **(C)**. Colors correspond to excitatory-enriched (blue) and inhibitory-enriched (red) genes identified through differential expression analysis of excitatory and inhibitory neurons.

(F) Pearson correlation of excitatory- and inhibitory-specific RNA-seq libraries depicted in **(E)**. Data shown is representative of 4 biological replicates. Note lower pearson correlation and clear dispersal of cell type-specific genes.

(G) Browser snapshot of *Dlgap1* genomic locus in both excitatory and inhibitory neurons of TAVI male mice at 6 weeks of age. Both RefSeq and Ensembl gene annotations are shown.

*P < 0.5, **P < 0.01, ***P < 0.001, ****P < 0.0001, n.s. = not significant; All data depicts mean ± SEM. See also **Figure 3.4**.

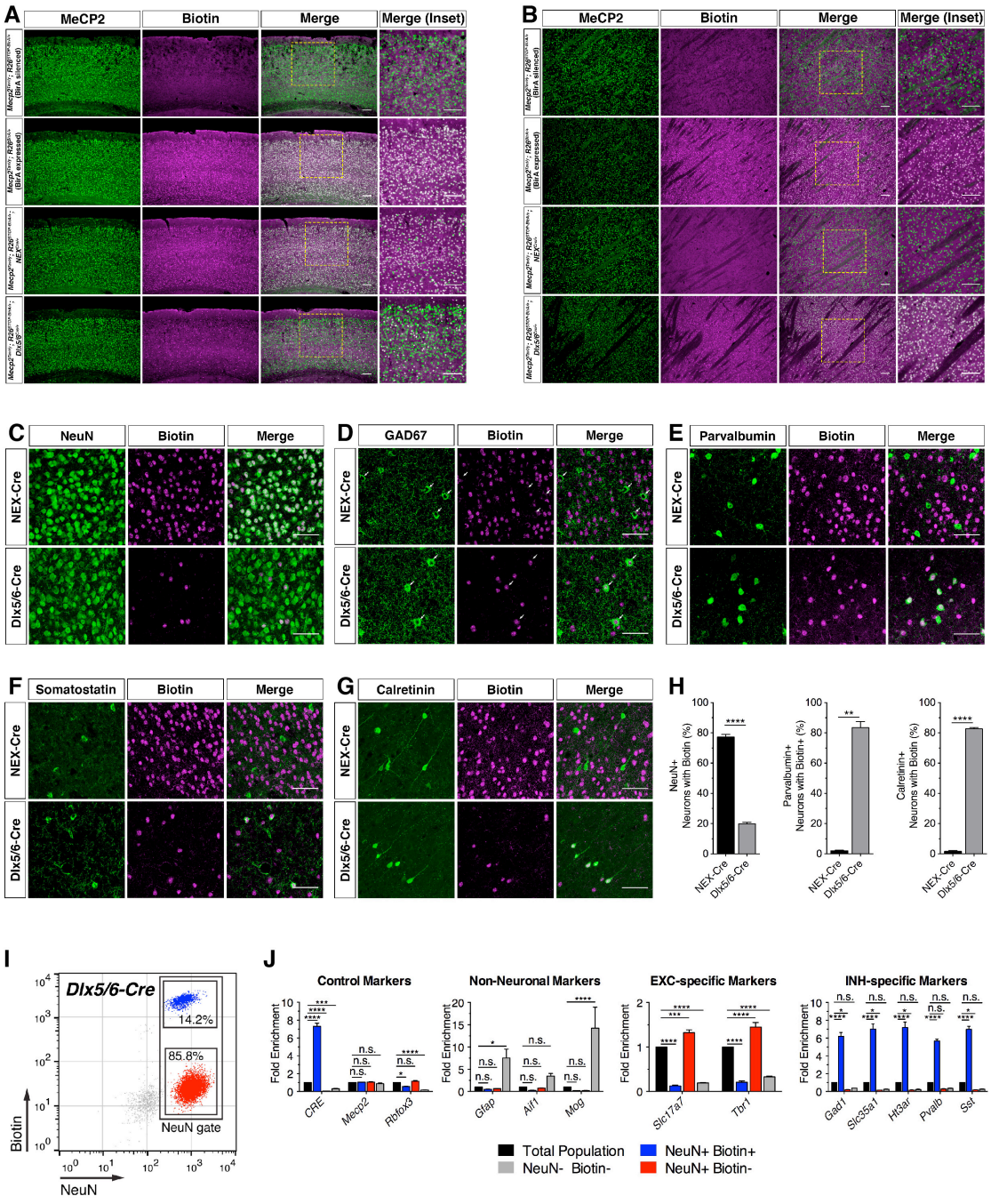

FIGURE 3.4: Validation of MeCP2-Tavi biotinylation in different cell types

(A,B) Representative immunofluorescent detection of biotinylated MeCP2-Tavi protein within Cre-specified neuronal populations of the **(A)** mouse cortex and **(B)** mouse striatum using an antibody against the MeCP2 C-terminus and streptavidin for biotin detection. Scale bars represent 100μm.

(C) Representative confocal images depicting Cre-mediated biotinylation in NeuN+ neurons. Scale bars represent 50μm.

(D) Representative confocal images depicting Cre-mediated biotinylation in Gad67+ interneurons. Scale bars represent 50μm.

(E) Representative confocal images depicting Cre-mediated biotinylation in Parvalbumin+ interneurons. Scale bars represent 50μm.

(F) Representative confocal images depicting Cre-mediated biotinylation in Somatostatin+ interneurons. Scale bars represent 50μm.

(G) Representative confocal images depicting Cre-mediated biotinylation in Calretinin+ interneurons. Scale bars represent 50μm.

(H) Quantification of Cre-mediated biotinylatin in NeuN+, Parvalbumin+, or Calretinin+ neurons (n = 3, Student's t-test).

(I) FACS analysis of labeled cortical nuclei populations. Data shown is representative of three independent experiments using Dlx5/6-Cre mice. Percentages indicate the mean distribution of neurons that are biotin+ (inhibitory; 14.2% ± 0.4) or biotin- (excitatory; 85.8% ± 0.4).

(I) RT-PCR validation of FACS-isolated populations depicted in **(C)**.

**P* < 0.5, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n.s. = not significant. All data depicts mean ± SEM.

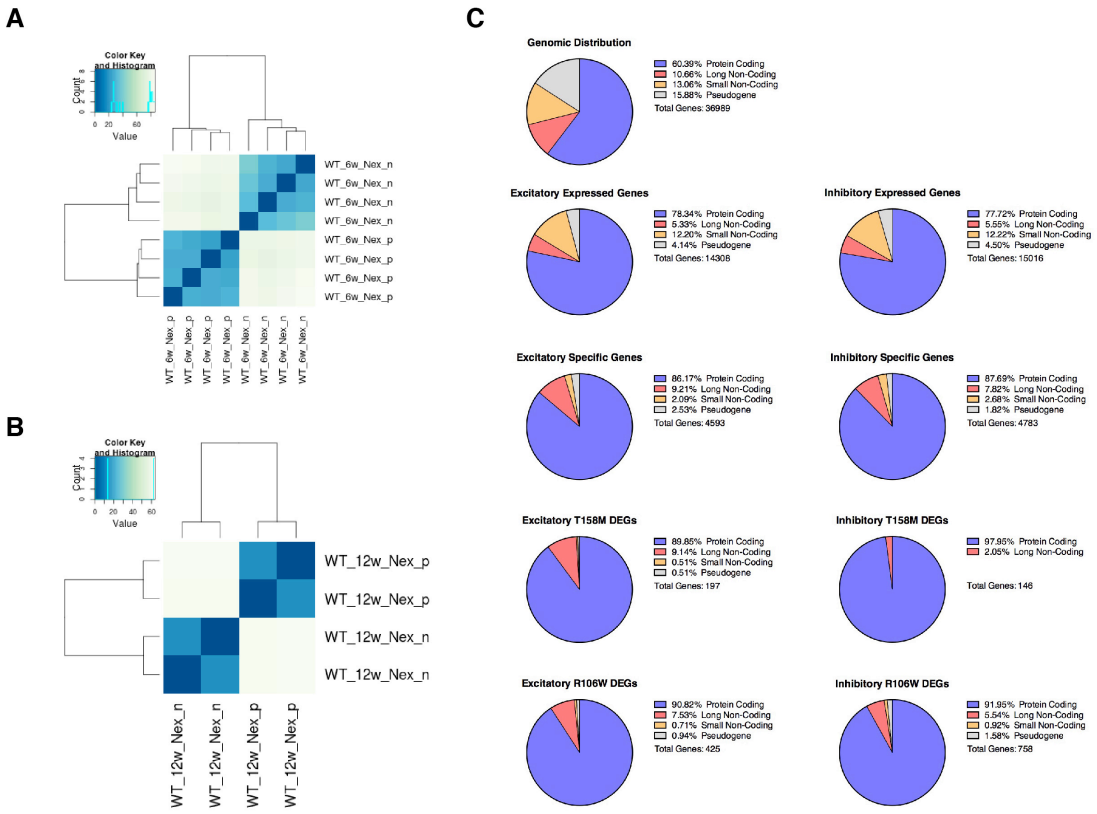

FIGURE 3.5: Validation of cell types by RNA-seq, and gene class distribution among DEGs

(A) Unsupervised hierarchical clustering and heatmap depicting correlations between 6 week excitatory and inhibitory neuronal nuclei from TAVI mice.

(B) Unsupervised hierarchical clustering and heatmap depicting correlations between 12 week excitatory and inhibitory neuronal nuclei from TAVI mice.

(C) Distribution of coding and non-coding genes differentially expressed genes (DEGs) in excitatory and inhibitory neurons of T158M and R106W male mice, compared to genomic, expressed, or cell type-enriched genes in each cell type.

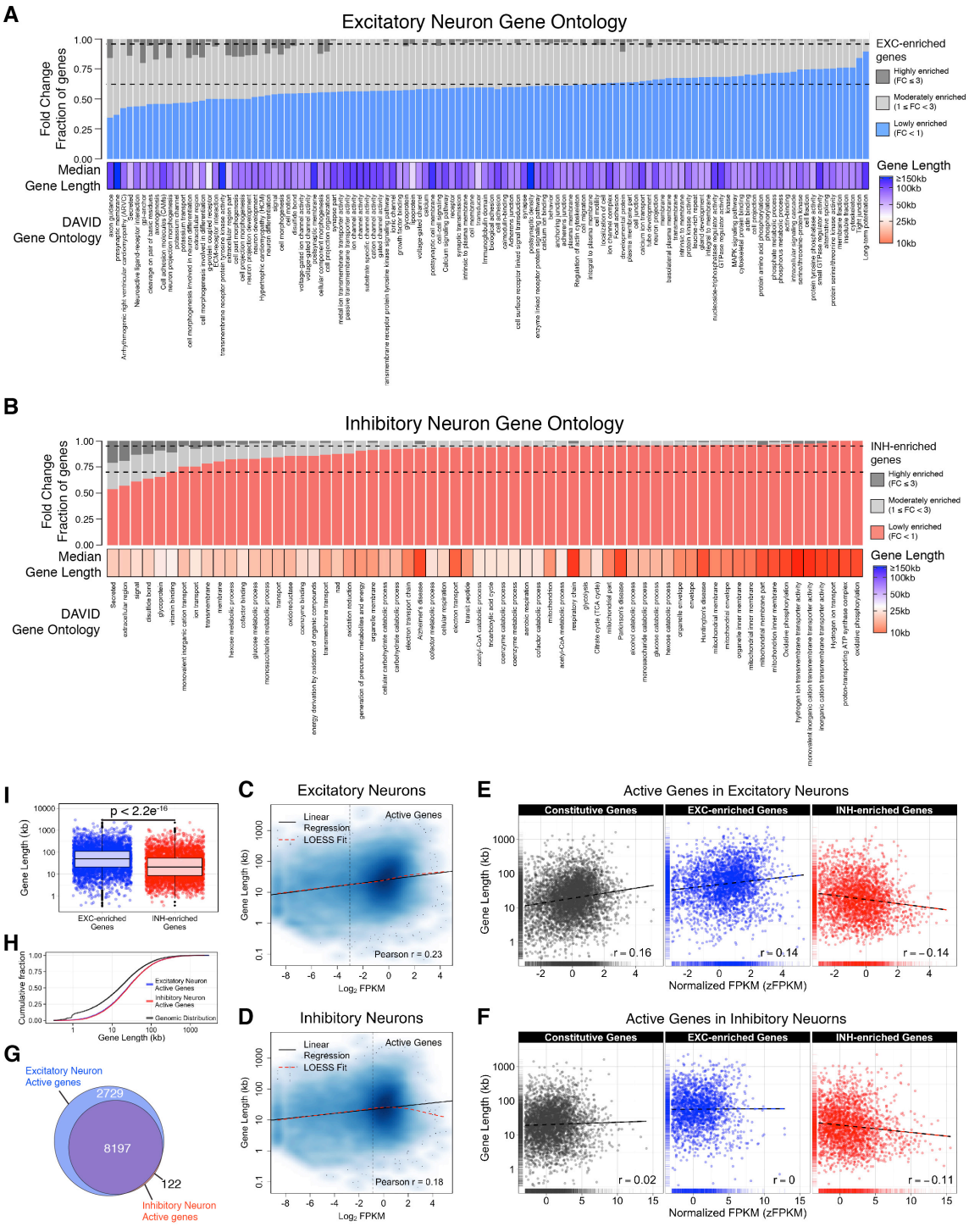

FIGURE 3.6: Characterization of EXC and INH neuronal gene expression

(A) Gene ontology for excitatory-enriched genes and their median gene lengths, sorted by level of enrichment (low, moderate, high) relative to inhibitory neurons.

(B) Gene ontology for inhibitory-enriched genes and their median gene lengths, sorted by level of enrichment (low, moderate, high) relative to excitatory neurons.

(C) Boxplot depicting gene lengths of excitatory- and inhibitory-enriched genes (Wilcoxon Rank Sum test)

(D) Cumulative distribution function of excitatory-enriched and inhibitory-enriched gene lengths ($P < 2.2e^{-16}$, Kolmogorov-Smirnov test)

(E) Area-proportional Venn diagram depicting proportion of active expressed genes in excitatory and inhibitory neurons that overlap.

(F) Smoothed scatterplot depicting correlation between gene length and $\log_2$ FPKM for all protein-coding genes in excitatory neurons (Pearson r = 0.23).

(G) Smoothed scatterplot depicting correlation between gene length and $\log_2$ FPKM for all protein-coding genes in excitatory neurons (Pearson r = 0.18).

(H) Scatterplot depicting correlation between gene length and $\log_2$ FPKM among all active expressed genes in excitatory neurons (depicted in C), parsed by constitutive (black), excitatory- (blue), or inhibitory-enriched genes (red).

(I) Scatterplot depicting correlation between gene length and $\log_2$ FPKM among all active expressed genes in inhibitory neurons (depicted in C), parsed by constitutive (black), excitatory- (blue), or inhibitory-enriched genes (red).

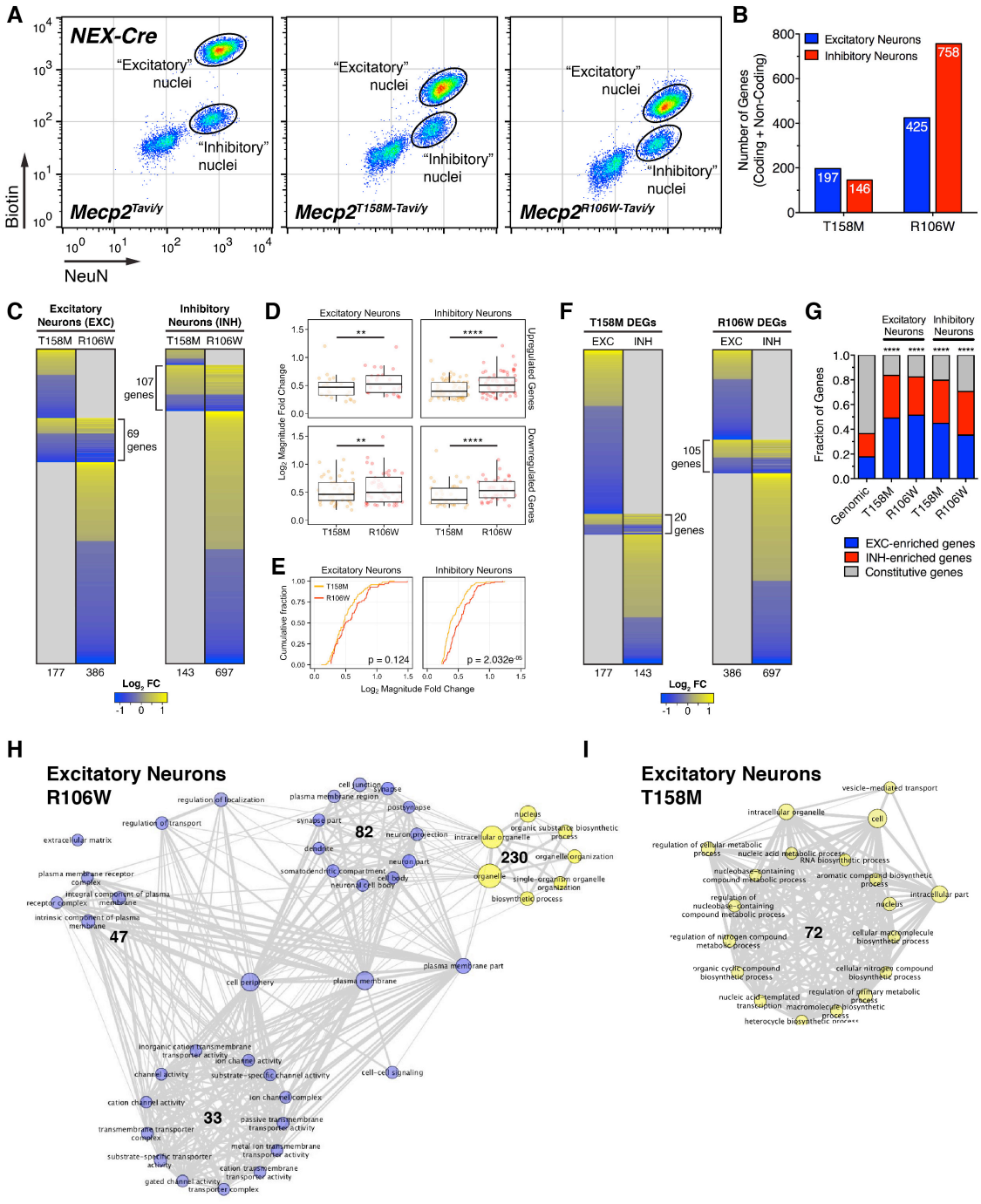

FIGURE 3.7: T158M and R106W differentially expressed genes at 6 weeks of age

(A) FACS isolation of cortical excitatory and inhibitory neuronal nuclei from hemizygous mice bearing the tagged wild-type (*Mecp2*$^{Tavi/y}$) or RTT mutant allele (*Mecp2*$^{T158M\text{-}Tavi/y}$, *Mecp2*$^{R106W\text{-}Tavi/y}$).

(B) Total number of differentially expressed genes (DEGs) identified within cortical excitatory or inhibitory neurons of RTT mutant mice.

(C) Heatmap displaying gene expression $\log_2$ fold changes within excitatory and inhibitory neurons of T158M and R106W mutant mice (comparison across genotype).

(D) Boxplots comparing the median fold changes between overlapping genes of T158M and R106W neurons depicted in **(C)**.

(E) Cumulative distribution function of the magnitude fold changes between T158M and R106W overlapping genes depicted in **(C)**.

(F) Heatmap displaying gene expression $\log_2$ fold changes within excitatory and inhibitory neurons of T158M and R106W mutant mice (comparison across cell type).

(G) Fraction of protein-coding DEGs that are also classified as constitutive, excitatory- or inhibitory-enriched genes across two independent RTT mouse models (Chi-square test)

(H) Enrichment map of gene ontology associations using pre-ranked Gene Set Enrichment Analysis (GSEA) with differentially expressed genes from R106W excitatory neurons ($P < 0.01$, FDR < 0.1). Blue represents gene sets associated with downregulated genes. Yellow represents gene sets associated with upregulated genes. Line weight denotes the extent of gene overlap between connected nodes.

(I) Enrichment map of gene ontology associations using pre-ranked GSEA with differentially expressed genes from T158M excitatory neurons ($P < 0.01$, FDR < 0.1). Blue represents gene sets associated with downregulated genes. Yellow represents gene sets associated with upregulated genes. Line weight denotes the extent of gene overlap between connected nodes.

*$P < 0.5$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, n.s. = not significant. See also **Figure 3.8**.

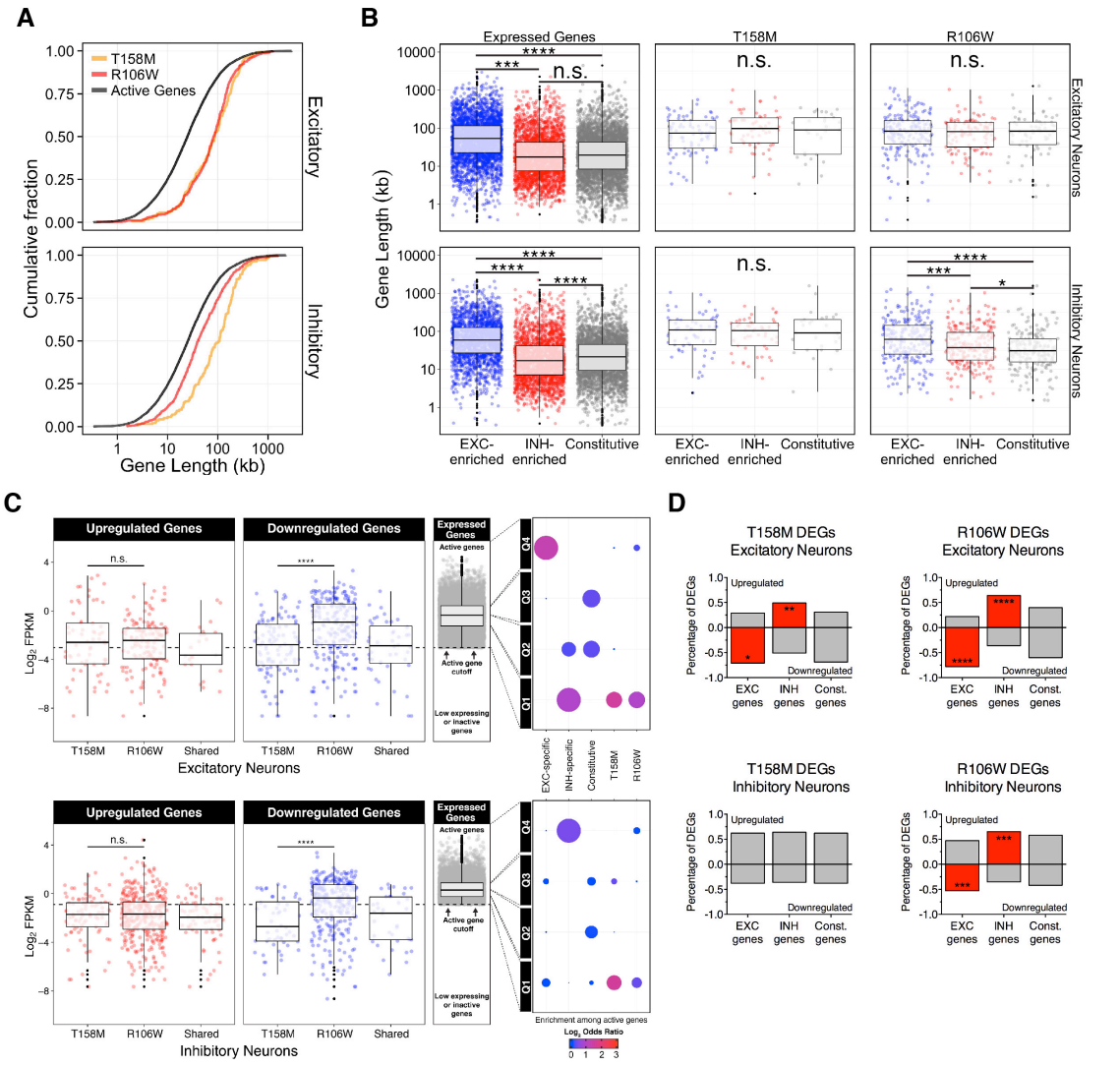

FIGURE 3.8: Additional expression features of T158M and R106W DEGs

(A) Cumulative distribution functions of protein-coding gene lengths between T158M and R106W male DEGs compared to the active expressed genes in their respective cell type.

(B) Box plots depicting protein-coding gene lengths among T158M and R106W male DEGs compared to the active expressed genes in their respective cell type, parsed by constitutive (black), excitatory- (blue), or inhibitory-enriched genes (red).

(C) *Left*, Boxplots comparing $\log_2$ FPKM distributions between T158M and R106W neurons. Additional boxplot shows distribution of shared genes between genotypes (immediately adjacent) and active expressed genes for each cell type. *Right*, Bubble plot depicting significant enrichment of DEGs across different expression quartiles as determined through Fisher's Exact Test. Area of circle scales with statistical significance. Q1 represents top 25% of active expressed genes (high expression), Q4 represents bottom 25% of active expressed genes (low expression).

(D) Bar graphs depicting percentage of DEGs that are preferentially upregulated or downregulated for each genotype/cell type. Statistically significant associations are depicted in red (Fisher's Exact Test).

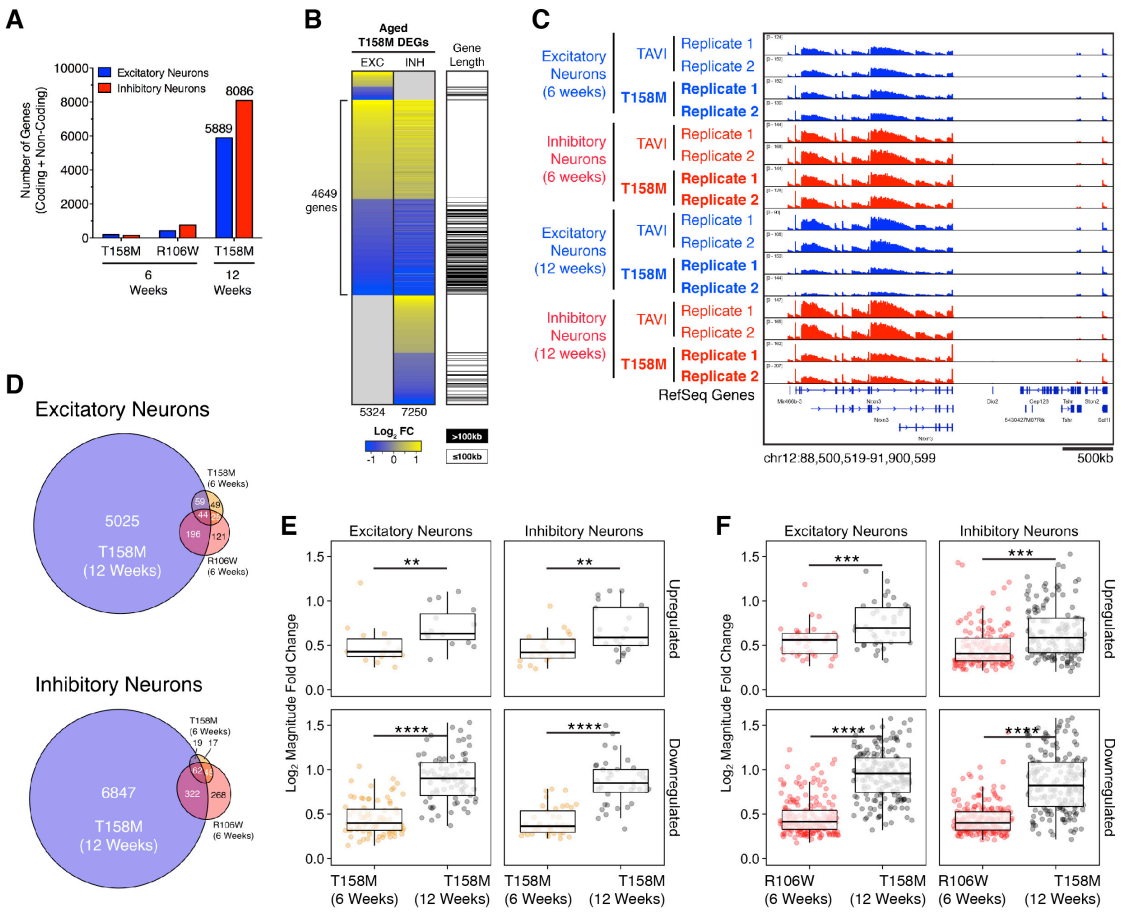

FIGURE 3.9: T158M differentially expressed genes at 12 weeks of age

(A) Total number of differentially expressed genes (DEGs) identified within cortical excitatory or inhibitory neurons of RTT mutant mice.

(B) Heatmap displaying gene expression $\log_2$ fold changes within excitatory and inhibitory neurons of T158M and R106W mutant mice (comparison across cell type, left) and associated gene lengths (right).

(C) Browser snapshot of *Nrxn3* genomic locus in both excitatory and inhibitory neurons of TAVI and T158M male mice at 6 and 12 weeks of age. Note reduction in intron-mapped reads at the *Nrxn3* gene among T158M neurons at 12 weeks of age, which are absent at 6 weeks of age.

(D) Area-proportional Venn diagram comparing MeCP2 DEG overlap between aged T158M, young T158M, and young R106W mice for excitatory (top) and inhibitory neurons (bottom).

(E) Boxplots comparing the median fold changes between overlapping genes in young and aged T158M neurons depicted in **(D)** (Wilcoxon Signed Rank test).

(F) Boxplots comparing the median fold changes between overlapping genes in young R106W and aged T158M neurons depicted in **(D)** (Wilcoxon Signed Rank test).

*P < 0.5, **P < 0.01, ***P < 0.001, ****P < 0.0001, n.s. = not significant. See also **Figure 3.11**.

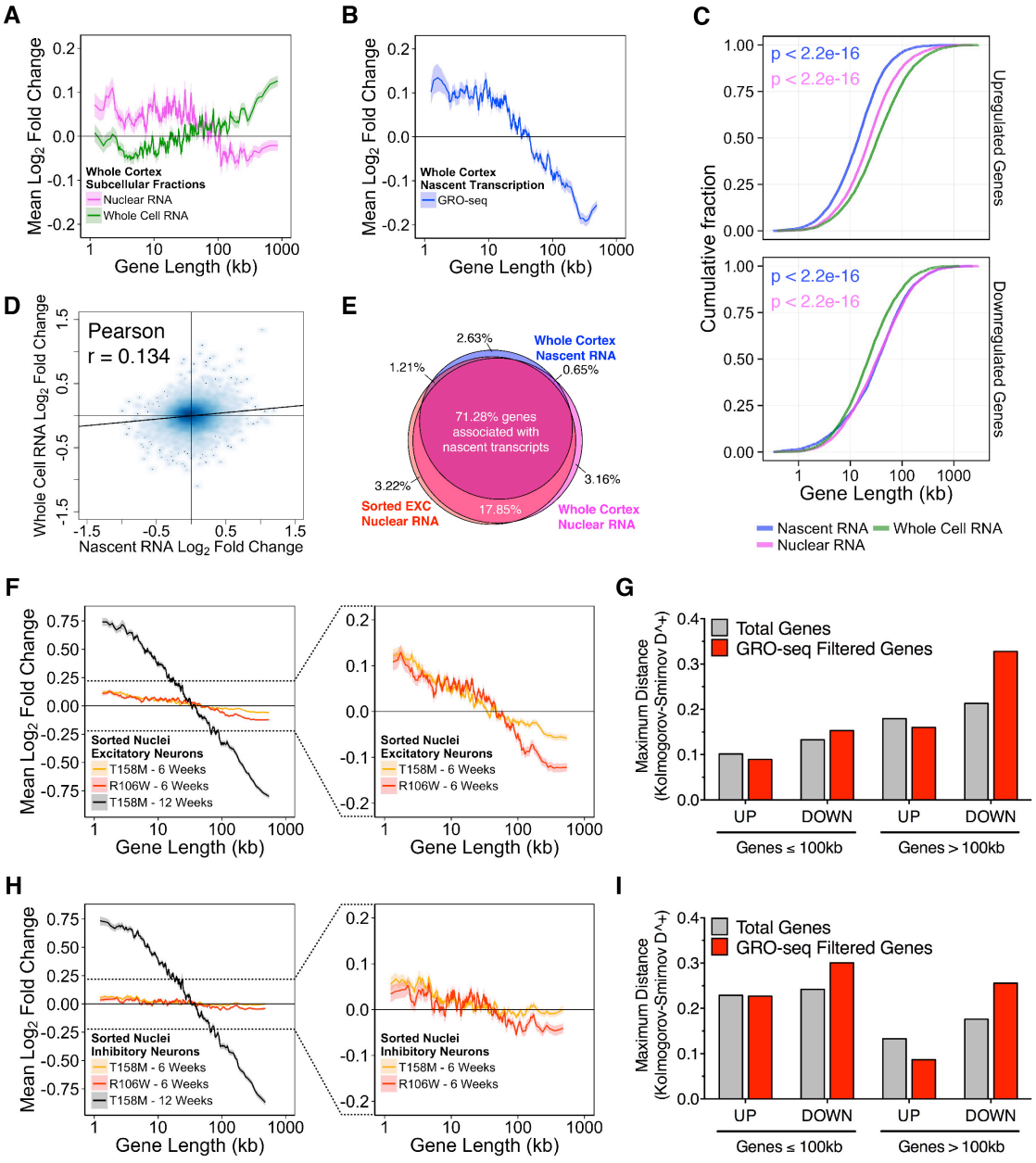

FIGURE 3.10: Genome-wide length-dependent transcriptional changes in RTT mutant mice

(A) Genome-wide changes in gene expression of R106W mice (n = 2) compared to TAVI mice (n = 2) at 6 weeks of age, as assessed by RNA-seq analysis of whole cell (green) or nuclear (magenta) RNA extracted from the same cortical tissue samples. Lines represent mean fold change in expression for genes binned according to gene length (200 gene bins, 40 gene step) as described in (Gabel et al., 2015). Ribbon represents the SEM of genes in each bin.

(B) Same as in **(A)**, but using nascent RNA from global nuclear run-on sequencing (GRO-seq).

(C) Cumulative distribution function of gene lengths for all upregulated and downregulated coding genes used in **(A)** and **(B)** among nascent (n = 10,951 genes), nuclear (n = 13,400 genes), and whole cell (n = 13,400 genes) RNA fractions. (P < 2.2e-16 for each nascent or nuclear RNA versus whole cell RNA comparison, Kolmogorov-Smirnov test).

(D) Scatterplot comparing $\log_2$ fold changes between nascent and whole cell RNA fractions.

(E) Area-proportional Venn diagram revealing proportions of genome-wide gene expression changes in nuclear RNA fractions (sorted excitatory neurons, whole cortical tissue) that are also transcriptionally engaged with RNA polymerase given their overlap with nascent RNA fractions.

(F) Mean fold change in gene expression of sorted excitatory neurons from young R106W (red; n = 4), young T158M (orange, n = 4) and aged T158M mice (n = 2) using transcriptionally engaged genes binned according to gene length.

(G) Genome-wide Kolmogorov-Smirnov maximum distance (D^+) between cumulative distributions of $\log_2$ fold changes among young T158M and R106W sorted excitatory neurons. See **Figure 3.11** for more details.

(H) Same as in **(F)**, but using sorted inhibitory neurons.

(I) Same as in **(G)**, but using sorted inhibitory neurons. See **Figure 3.11** for more details.

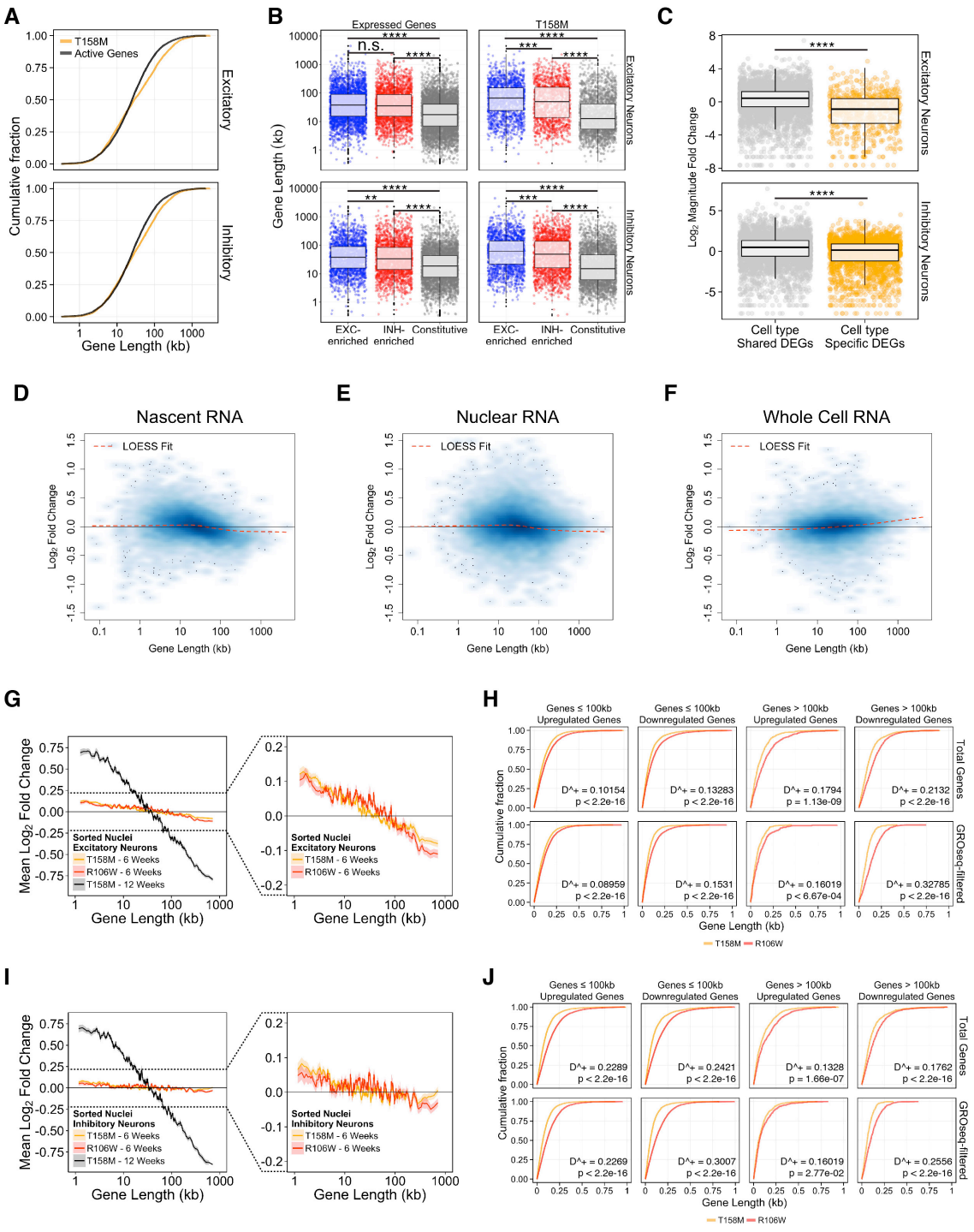

FIGURE 3.11: Genome-wide length-dependent transcriptional changes in RTT mutant mice (Extended)

(A) Cumulative distribution functions of protein-coding gene lengths among aged T158M male DEGs compared to the active expressed genes in their respective cell type.

(B) Box plots depicting protein-coding gene lengths among aged T158M male DEGs compared to the active expressed genes in their respective cell type, parsed by constitutive (black), excitatory- (blue), or inhibitory-enriched genes (red).

(C) Boxplots comparing $\log_2$ FPKM distributions (expression levels) of 12 week T158M DEGs that are either shared between or specific to excitatory and inhibitory neurons (Wilcox Rank Sum Test).

(D) Smoothed scatterplot depicting correlation between gene length and $\log_2$ fold change for all protein-coding genes from nascent RNA fractions (Whole Cortex GRO-seq).

(E) Smoothed scatterplot depicting correlation between gene length and $\log_2$ fold change for all protein-coding genes from nuclear RNA fractions (Whole Cortex nuclear RNA-seq).

(F) Smoothed scatterplot depicting correlation between gene length and $\log_2$ fold change for all protein-coding genes from whole cell RNA fractions (Whole Cortex RNA-seq).

(G) Mean fold change in gene expression of sorted excitatory neurons from young R106W (red; n = 4), young T158M (orange, n = 4) and aged T158M mice (n = 2). Not sorted for transcriptionally engaged (GRO-seq detected) genes.

(H) Genome-wide Kolmogorov-Smirnov maximum distance (D^+) between cumulative distributions of $\log_2$ fold changes among young T158M and R106W sorted excitatory neurons.

(I) Mean fold change in gene expression of sorted inhibitory neurons from young R106W (red; n = 4), young T158M (orange, n = 4) and aged T158M mice (n = 2). Not sorted for transcriptionally-engaged (GRO-seq detected) genes.

(J) Genome-wide Kolmogorov-Smirnov maximum distance (D^+) between cumulative distributions of $\log_2$ fold changes among young T158M and R106W sorted inhibitory neurons.

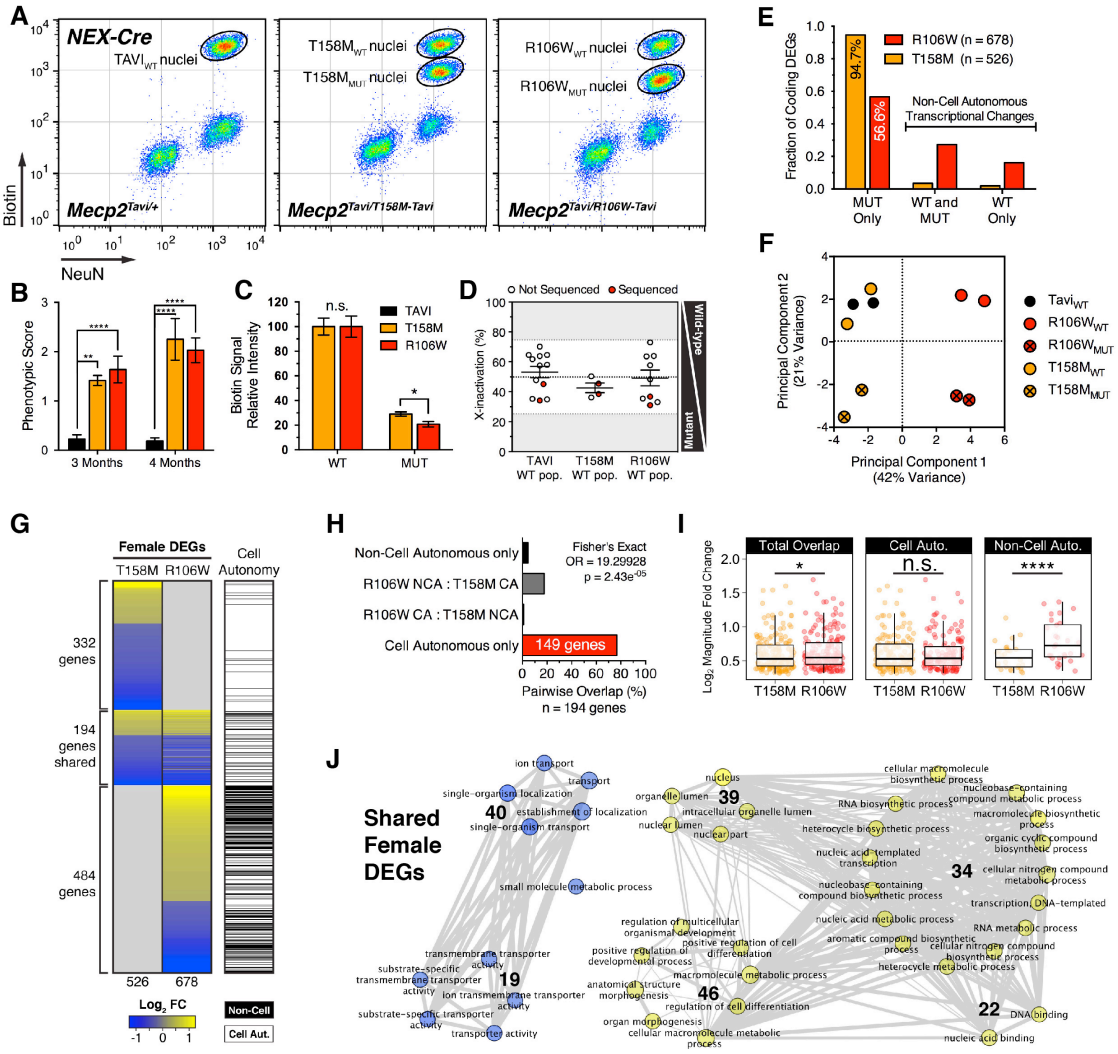

FIGURE 3.12: Transcriptional profiling of genetically mosaic neurons in RTT female mice

(A) FACS isolation of genetically-mosaic cortical excitatory neuronal nuclei from heterozygous female mice bearing the tagged wild-type (*Mecp2*$^{Tavi/+}$) or RTT mutant allele (*Mecp2*$^{Tavi/T158M\text{-}Tavi}$, *Mecp2*$^{Tavi/R106W\text{-}Tavi}$).

(B) RTT-like phenotypic presentation over postnatal age in tagged TAVI (n = 12), T158M (n = 4), and R106W (n = 9) female mice (F = 41.14 (Genotype), *P < 0.0001*, two way ANOVA).

(C) Quantification of biotin signal from FACS-isolated populations depicted in **(A)** (n_{T158M} = 4, n_{R106W} = 9, Student t-test).

(D) Distribution of X-inactivation among cortical excitatory neurons from FACS-isolated nuclei as a percentage of the WT population. Data points in red highlight samples that were used for high-throughput sequencing.

(E) Bar graph depicting the total distribution of unique cell and non-cell autonomous DEGs between T158M and R106W female excitatory neurons.

(F) Principal components analysis of WT and MUT populations isolated from TAVI, T158M, and R106W female mice.

(G) *Left,* Heatmap displaying the total number of unique gene expression $\log_2$ fold changes detected from both WT and MUT populations in either T158M or R106W female excitatory neurons. Note genes that overlap across genotypes (n = 194). *Right*, Heatmap displaying the distribution of cell (Cell Aut.) and non-cell autonomous (Non-Cell) DEGs.

(H) Proportion of cell autonomous and non-cell autonomous genes that overlap between T158M and R106W female excitatory neurons.

(I) Boxplots comparing the median fold changes between total, cell autonomous, and non-cell autonomous overlapping genes between T158M and R106W female excitatory neurons. (one-tailed Wilcoxon Signed Rank test)

(J) Enrichment map of gene ontology associations using pre-ranked Gene Set Enrichment Analysis (GSEA) with shared DEGs between T158M and R106W female excitatory neurons (P < 0.01, FDR < 0.1). Blue represents gene sets associated with downregulated genes. Yellow represents gene sets associated with upregulated genes. Line weight denotes the extent of gene overlap between connected nodes.

*P < 0.5, **P < 0.01, ***P < 0.001, ****P < 0.0001, n.s. = not significant. See also **Figure 3.13**.

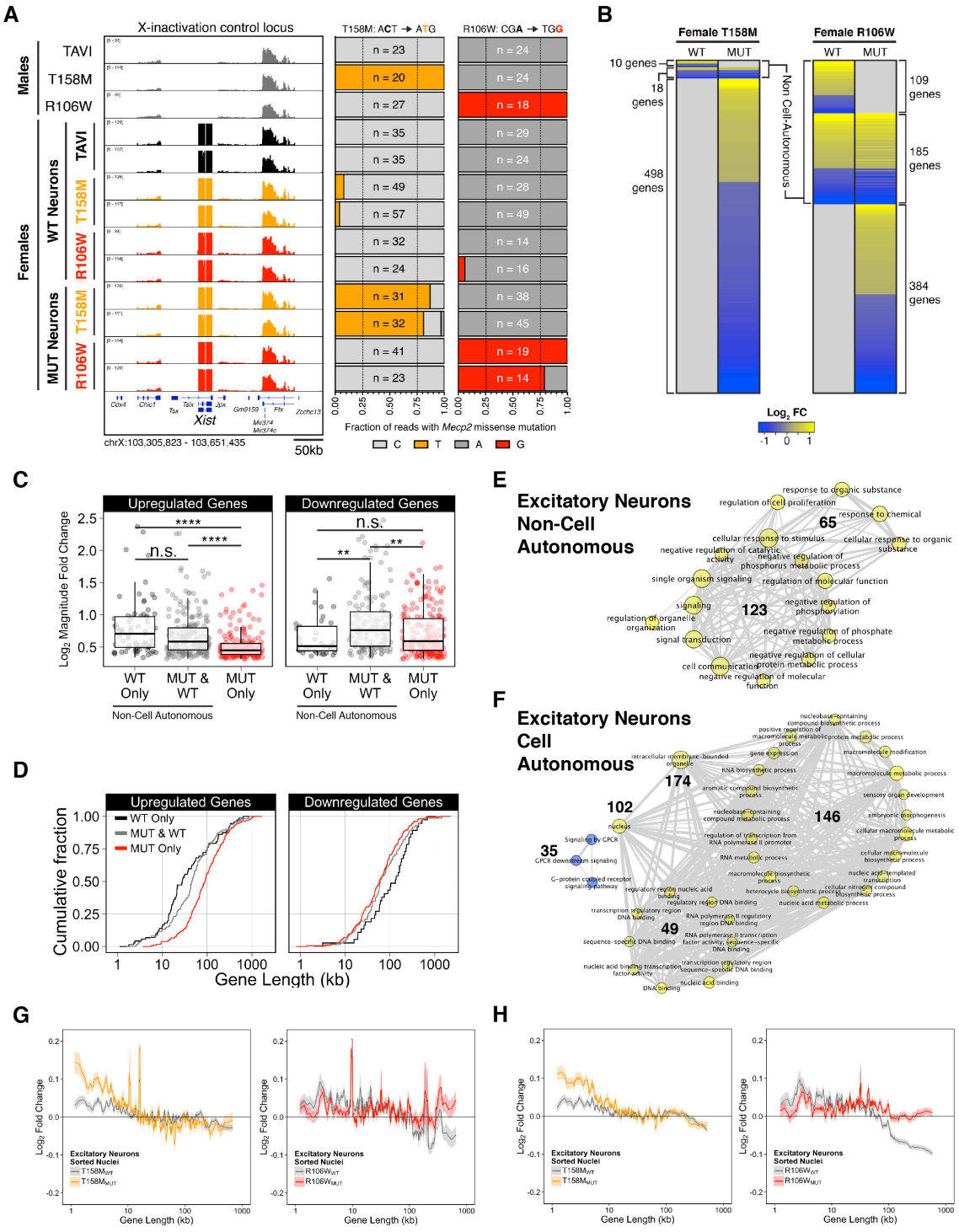

FIGURE 3.13: Cell and non-cell autonomous transcriptional changes are molecularly distinct processes

(A) *Left*, Browser snapshot of X-inactivation control genomic locus in both WT and MUT neurons of TAVI, T158M, and R106W female mice at 18 weeks of age. *Right*, Percentage of reads detected with wild-type or knockin allele for each sorted population.

(B) Heatmap displaying the total gene expression $\log_2$ fold changes detected in both WT and MUT populations of T158M or R106W female excitatory neurons.

(C) Boxplot comparing the magnitude of $\log_2$ fold change among cell and non-cell autonomous genes (Wilcoxon Rank Sum test).

(D) Cumulative distribution function comparing the gene lengths among cell and non-cell autonomous genes (Wilcoxon Rank Sum test).

(E) Enrichment map of gene ontology associations using pre-ranked Gene Set Enrichment Analysis (GSEA) with non-cell autonomous DEGs in R106W female excitatory neurons ($P < 0.01$, FDR < 0.1). Blue represents gene sets associated with downregulated genes. Yellow represents gene sets associated with upregulated genes. Line weight denotes the extent of gene overlap between connected nodes.

(F) Enrichment map of gene ontology associations using pre-ranked Gene Set Enrichment Analysis (GSEA) with cell autonomous DEGs in R106W female excitatory neurons ($P < 0.01$, FDR < 0.1). Blue represents gene sets associated with downregulated genes. Yellow represents gene sets associated with upregulated genes. Line weight denotes the extent of gene overlap between connected nodes.

(G) Mean fold change in gene expression of sorted WT or MUT excitatory neurons from T158M (orange; n = 2) or R106W (red, n = 2). Not sorted for transcriptionally-engaged (GRO-seq detected) genes.

(H) Mean fold change in gene expression of sorted WT or MUT excitatory neurons from T158M (orange; n = 2) or R106W (red, n = 2) mice using transcriptionally engaged genes binned according to gene length.

*$P < 0.5$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, n.s. = not significant.

CHAPTER 4: Discussion, Implications, and Future Directions

4 General Summary of Dissertation Work

Despite two decades of strong genetic and experimental evidence linking MeCP2 to RTT and associated clinical symptoms and cellular phenotypes, the mechanistic function of this protein still remains unclear. Many studies have attempted to ascribe a function through MeCP2-dependent gene expression changes by using whole tissue mRNA extracted from *Mecp2*-null male mice for profiling. However, the limited numbers of consistent gene expression changes across multiple studies and the subtle degree to which these genes are misregulated have hindered efforts to define a consistent molecular role for MeCP2 in RTT. Recent studies are beginning to illuminate the extensive heterogeneity and complexity of gene expression and regulation at the cellular and subcellular level, especially in the brain. This diversity, coupled with the clinical variability of RTT-associated mutations among patients that are predominantly female, argues for a more context-specific experimental approach towards investigating MeCP2 function that is only now being appreciated.

In Chapter 2, I began to address these issues of cellular heterogeneity by profiling gene expression in *Mecp2*-null mice within the striatum, a brain region whose neurons comprise a single cell type (>95% medium spiny neurons). By comparing gene expression changes in the striatum to additional neuronal (hypothalamus, cerebellum) and non-neuronal (liver) tissues, I largely demonstrated that most expression changes are specific to individual brain regions, although the affected genes are typically expressed throughout the brain, and questioned whether this would apply to individual cell types as well.

In Chapter 3, I significantly expanded upon the previous study by developing a novel genetic strategy that allows for cell type-specific transcriptional profiling and investigation of MeCP2 function across multiple neuronal contexts. I also developed an allelic series of *Mecp2*-mutant mice bearing RTT-associated missense mutations. By combining Cre-mediated biotin

tagging of MeCP2 and RTT-associated mutant variants *in vivo* with RNA sequencing from FACS-purified neuronal nuclei of adult mice, I was able to 1) identify transcriptional changes in cortical excitatory and inhibitory neurons, 2) investigate how these changes scale with moderate to severe impairments in binding to methylated DNA, 3) observe how these changes relate to neuronal gene expression in a physiological context, and 4) distinguish molecularly distinct cell autonomous and non-cell autonomous transcriptional effects within genetically mosaic female mice. Taken together, these findings allow me to explore a speculative model whereby transcriptional deficits in the absence of MeCP2 binding lead to the neuronal and circuit level deficits observed in RTT patients and *Mecp2* animal models, with mechanistic implications for MeCP2 in modulating neuronal plasticity in a physiological context.

4.1 MBD MISSENSE MUTATIONS MODEL RETT SYNDROME IN MICE

4.1.1 Summary of Findings

I developed knockin mice bearing RTT-associated missense mutations that either partially (T158M) or completely abolish (R106W) MeCP2's ability to bind to methylated DNA. These mice closely resembled *Mecp2*-null mice in the severity and onset of most gross phenotypes, consistent with the increased phenotypic severity of mutations in the MBD (Cuddapah et al., 2014). However, R106W mice closely resembled *Mecp2*-null mice in terms of lifespan, whereas the T158M mice were more mildly affected and survived an additional 4 weeks; thus these mice phenotypically correlate with the degree of molecular impairment.

4.1.1 Broader Significance

Both typical and atypical RTT patients bearing these two mutations also exhibit a similar relationship to what I observed among mice, and are associated with intermediate (T158M) and severe (R106W) clinical diagnostic scores. Experimental studies comparing the phenotypic effects of RTT mutations further corroborate these observations on a molecular scale. Biochemically, the MBD is the most structured domain of MeCP2. Interdomain interactions, which include the ASX-ST motif, help stabilize MBD structure and are required for forming and

maintaining hydrogen bond interactions with methyl groups on DNA (Ghosh et al., 2008; Ho et al., 2008). Both mutations are predicted to introduce hydrophobic charges that destabilize (T158M) or completely disrupt (R106W) the local topology and hydrophilic environment of the ASX-ST motif, contributing to reduced affinity for methylated DNA. EMSA studies report that the R106W mutation consistently abolishes affinity for methylated DNA, whereas the T158M mutation exhibits partial to severe reductions in affinity relative to wild-type protein *in vitro* (Ballestar et al., 2000; Brown et al., 2015; Ghosh et al., 2008). The R106W mutation in particular also disrupts MeCP2's ability to compact nucleosomes, mediate higher-order structures of nucleosomal arrays, and confer MNase resistance *in vitro* (Nikitina et al., 2007b). Molecular FRAP studies also report an increase in the nuclear mobility and diffusion kinetics of MeCP2 bearing these two mutations, and this occurs to a greater extent with the R106W than the T158M mutation, indicative of reduced association with chromatin *in vivo* (Kumar et al., 2008b; Schmiedeberg et al., 2009). Finally, transcriptional repressor assays using methylated reporter constructs report a partial loss of repressive activity when *Drosophila S2* cells are transfected with MeCP2 T158M in contrast to MeCP2 R106W, which displays a complete loss of activity concomitant with a complete loss of methyl-CpG binding (Kudo et al., 2001).

4.2 EXCITATORY AND INHIBITORY NEURONS DISPLAY CELL TYPE-SPECIFIC TRANSCRIPTIONAL DYNAMICS

4.2.1 Summary of Findings

The genetic strategies that I developed to both separate and profile nuclear gene expression in cortical excitatory and inhibitory neurons allowed me to 1) identify genes whose expression is enriched in a particular cell type over the other, and 2) determine how neuronal genes, and particularly excitatory- and inhibitory-enriched genes, are typically expressed within an unperturbed physiological context. Previous studies have demonstrated that neuronally-expressed genes tend to be longer in gene length relative to genes expressed in other somatic tissues and cell types (Gabel et al., 2015; King et al., 2013; Zylka et al., 2015). I confirmed as

such among genes expressed in excitatory and inhibitory cortical neurons of adult mice, which were significantly longer than genomic distributions of gene length. The same was true for excitatory- and inhibitory-enriched genes, which tended to be cumulatively longer and shorter in gene lengths, respectively.

However, I observed additional cell-specific behaviors among active expressed genes in each cell type that have not been previously characterized. Excitatory and inhibitory neurons largely express the same group of genes. These include genes whose expression is lowly or moderately enriched ($\log_2$ FC < 3) in one cell type as opposed to the other. The manner in which these genes are expressed, however, is dramatically different. Short inhibitory-enriched genes, which are associated with cellular respiration and energy metabolism, are expressed in a length-dependent manner in both excitatory and inhibitory neurons, which would be expected of metabolically active neurons regardless of subtype. It is likely that these genes are more highly expressed in inhibitory neurons because, as mitochondria-rich cell types, they have perpetual role in dynamically suppressing the activity of excitatory neurons to maintain overall network homeostasis in the brain (Kann et al., 2014; Maffei and Fontanini, 2009). In contrast, long excitatory-enriched genes are positively correlated with the level of gene expression exclusively within excitatory neurons; this relationship is completely absent in inhibitory neurons. Excitatory-enriched genes encode proteins that promote axogenesis (signaling molecules, actin-associated proteins), are localized to the plasma membrane (synaptic proteins, ion channels, transmembrane receptors) and extracellular matrix (secreted and cell adhesion molecules, integrins), and participate in intracellular signaling cascades (protein kinases and phosphatases). Thus these genes are consistent with the role of cortical pyramidal neurons that undergo extensive changes in synaptic plasticity at the plasma membrane in response to neuronal circuit activity (Beaulieu, 1993; Emes et al., 2008; Holtmaat and Svoboda, 2009).

4.2.2 Broader Significance

Because excitatory- and inhibitory-enriched genes and their associated functions are consistent with the known roles of these two broad cortical cell types, I posit that excitatory

neurons exhibit a strong positive association between gene length and expression level in order to carry out an assortment of cell-specific functions required for proper neuronal activity in the brain. I would thus expect cortical pyramidal neurons to be particularly sensitive to factors that perturb this length-dependent transcriptional regulation, which is supported by the effects of DNA supercoiling and topoisomerase inhibition on mammalian cells (King et al., 2013; Kouzine et al., 2013; Mabb et al., 2014; Madabhushi et al., 2015; Naughton et al., 2013; Puc et al., 2015).

Topoisomerases are enzymes that relieve the torsional stress associated with DNA supercoiling, which occur as a consequence of both cell division and gene transcription (Chen et al., 2013; Tsao et al., 1989; Wang, 2002). Cultured cortical neurons treated with 300μm topotecan, a reversible TOP1 inhibitor, display impairments in excitatory synapse formation and function and deficits in inhibitory synapse function, leading to overall reduced spontaneous network activity without compromising cellular health. These effects are caused by an inability to resolve DNA supercoiling at particularly long genes, leading to impaired transcriptional elongation of RNA polymerase II and length-dependent downregulation of neuronal genes important for synaptic function, many of which include genes associated with autism. In addition to length-dependent transcription, topoisomerase 1 has also been shown to be required for ligand-dependent androgen receptor-mediated enhancer activation in prostate cancer cells by relieving torsional strain required for productive elongation of eRNAs (Puc et al., 2015). Immediate early gene activation and expression in depolarized primary cortical neurons also requires the production of double-stranded breaks by topoisomerase IIβ to relieve torsional strain and mediate synaptic changes (Madabhushi et al., 2015). Beyond the role of topoisomerases, these studies support the notion that high efficiency of transcriptional elongation may promote the expression of neuronal genes.

Notably, transcriptional elongation of genes can be impeded by well-ordered nucleosomes along the gene body, which increases the frequency of productively elongating Pol II pausing across the gene body in eukaryotic yeast cells (Churchman and Weissman, 2011). Increases in torsional stress can also destabilize well-order nucleosomes in *Drosophila* S2 cells

and lead to increased nucleosome turnover, which may impact overall chromatin structure at certain genes (Teves and Henikoff, 2014). Alternatively, high efficiency of transcriptional elongation may be promoted by participating in long-range chromatin interactions within transcriptionally active topologically associating domains (TADs) or other active sites of transcription in the nucleus (Dixon et al., 2012; Osborne et al., 2004). Transcriptional activity and elongation can also effect changes in large-scale chromatin structure by promoting large regions of positive or negative supercoiling of DNA that help to establish broad topological domains of condensed or decondensed chromatin, respectively (Naughton et al., 2013).

The complex interplay being chromatin structure and the transcriptional elongation of long genes in pyramidal neurons is particularly interesting in light of studies that detect significant genetic risk associations among chromatin modifiers, transcriptional regulators, and synaptic genes in autism, schizophrenia, and intellectual disability (Cotney et al., 2015; De Rubeis et al., 2014; Iossifov et al., 2014). I conclude that excitatory neurons promote the length-dependent expression of long genes required for overall synaptic function at the plasma membrane and may be particularly sensitive to chromatin-mediated barriers of efficient transcriptional elongation.

4.3 DOWNREGULATED GENES MODULATE RETT PHENOTYPES

4.3.1 Summary of Findings

Using aged-matched T158M and R106W mutant male mice at the onset of overt RTT phenotypes, I sequenced nuclear RNA from cortical excitatory and inhibitory neurons. Genome-wide, when MeCP2 function is generally impaired, the expression of long genes is predominantly downregulated in a length-dependent manner. This argues for MeCP2 in promoting the expression of long genes, regardless of cortical cell type. The degree to which long genes are downregulated correlates with the degree of MeCP2 impairment, which I defined—through the use of two RTT-associated missense mutations in the MBD—as MeCP2's ability to bind to methylated cytosines and, by proxy, chromatin *in vivo*. Furthermore, genome-wide differences between these two mutations were only detected among genes that can produce *de novo* RNA

transcripts from isolated cortical nuclei using GRO-seq, i.e. genes that are transcriptionally engaged with RNA Polymerase II. Thus, the active transcription of long genes in excitatory and inhibitory neurons is sensitive to differences in MeCP2's ability to bind to chromatin *in vivo*. Finally, the degree to which genes are downregulated worsens with age, which correlates with the progressive worsening of phenotypes when T158M mice are aged an additional 6 weeks.

Upon initiating this study, I hypothesized that MeCP2 DEGs (FDR < 0.05) would reveal genes that are most sensitive to the effects of MeCP2 binding. Expression features among FDR-controlled genes that correlate with MeCP2 impairment may thus provide refined insight into the types of genes that MeCP2 may mechanistically modulate. Consistent with genome-wide observations and my hypothesis, I found that aged-matched R106W mice exhibited more DEGs than T158M mice. Similarly, genes that are shared between both genotypes are more misregulated in R106W mice. The severity between these two mutations also correlated with the ability to discern functional enrichment from downregulated DEGs in age-matched mice (Gene Set Enrichment Analysis, FDR < 0.10). Functional terms could only be statistically inferred from the R106W genotype and suggested a mutation-specific loss of ion channel, glutamate receptor and synaptic function among excitatory neurons. Interestingly, these terms were also identified among downregulated genes in aged female T158M and R106W excitatory neurons. Taken together, observations from FDR-controlled DEGs suggest that the cell autonomous loss of synaptic genes that localize to the somatodendritic or post-synaptic compartment correlates with RTT phenotypic severity, which occurs in both male and female neurons and depends on MeCP2's ability to bind to methylated DNA. This is in particular agreement with genome-wide transcriptional observations in excitatory neurons, where the global downregulation of long genes would be expected to significantly impact synaptic functions (described in 4.2.1).

4.3.2 Broader Significance

Given the strong similarities with topoisomerase 1 inhibition and length-dependent transcriptional downregulation of long genes that lead to synaptic deficits, the transcriptional

effects that are observed in T158M and R106W neurons likely underlie the neuronal deficits observed in both RTT patients and mouse models, which include reduced dendritic branching, reduced number and function of glutamatergic synapses, and reduced hypoexcitability of excitatory pyramidal neurons. In the context of excitatory neurons, it is remarkable that the degree of MeCP2 impairment across mutations correlates with the degree of transcriptional deficits, which may ultimately modulate the degree of synaptic phenotypes and help explain the spectrum of clinical severity among RTT patient mutations. Differences in synaptic impairments across mutations would explain why the R106W female mice have non-cell autonomous gene expression changes in contrast to T158M female mice. Many of the genes associated with non-cell autonomous changes are neuroprotective and help ameliorate RTT-mediated neuronal and behavioral phenotypes in mice. Synaptic function in excitatory neurons, and overall neuronal network activity, may be more severely compromised in R106W mice than T158M mice, leading to the upregulation of genes that help to boost neuronal synaptic plasticity.

The differences in transcriptional deficits that correlate with MeCP2 impairment are likely due to the differences in the affinity of MeCP2 T158M and MeCP2 R106W for methylated DNA *in vivo*, rather than reduced protein expression levels. Experimental evidence in our lab supports the notion that MeCP2 is a stable protein when bound to chromatin, and that mutations that disrupt binding to chromatin (the primary effect) leads to increased protein instability and turnover (the secondary effect) (J. Lamonica, personal communications). Several knockin *Mecp2* mouse models bearing RTT-associated mutations in the MBD also show reduced protein stability (Brown et al., 2015; Goffin et al., 2012). When considered together with our knockin mice, MBD-associated mutations demonstrate a positive correlation between increased ability to bind to methylated DNA and increased protein stability, which lends support to our hypothesis. This correlation therefore establishes that MeCP2 protein expression levels among RTT-associated mutations in the MBD may serve as a general readout for residual binding to chromatin. Under this assumption, the T158M mutation has a higher residual binding capacity to chromatin than the R106W mutation, which I consistently observed in both male and female mice. This is in

agreement with molecular FRAP studies in live cells which demonstrate that, in contrast to wild-type MeCP2 residence time on chromatin (~15 seconds), MeCP2 T158M spends ~4 seconds on chromatin, whereas MeCP2 R106W spends < 2.5 seconds on chromatin (Schmiedeberg et al., 2009).

Given the importance of length-dependent transcription in excitatory neurons that I previously established (see section 4.2), I propose that MeCP2 promotes chromatin-mediated increases in the efficiency of transcriptional elongation by directly binding to long neuronal genes. This may be consistent with MeCP2 being chromatin-bound and localized within highly nuclease-accessible regions in the brain (Rube et al., 2016; Thambirajah et al., 2012). Bisulfite-sequencing of neurons do show that gene body methylation in both CpG and CpA contexts are found at neuronal genes, and MeCP2 ChIP-seq displays increased read density at long genes over methyl-CpG and methyl-CpA sites (Baubec et al., 2013; Chen et al., 2015; Gabel et al., 2015). The increase in MeCP2 expression mirrors the increase in methyl-CpA deposition as neurons mature and undergo synaptogenesis, and methyl-CpA density increases as genes increase in length (Chen et al., 2015; Gabel et al., 2015; Lister et al., 2013; Mo et al., 2015; Skene et al., 2010). Methyl-CpG density also increases as gene length increases, but to a much lesser extent than methyl-CpA, likely owing to its relatively higher abundance in the genome (Gabel et al., 2015). The high density of methyl-CpG and methyl-CpA over gene bodies may therefore lead to increased cooperative binding of MeCP2 to long neuronal genes (Baubec et al., 2013; Ghosh et al., 2010).

One method in which MeCP2 binding may promote increased transcriptional efficiency is through promoting more open and de-condensed chromatin structures and long genes. Such chromatin arrangements are more favorable for nucleosome sliding and remodeling, which is required to remove barriers for efficient transcriptional elongation (Bell et al., 2011; Churchman and Weissman, 2011; Clapier and Cairns, 2009; Zentner et al., 2013). This is supported by observations from a recent high-resolution *in vivo* imaging study of genetically mosaic neurons in *Mecp2*-null female mice (Linhoff et al., 2015). In this study, they found that *Mecp2*-null

hippocampal CA1 neurons display a 20% increase in chromatin compaction *in vivo*. This was concomitant with an increase in global H4K20me3, which correlates with transcriptional repression at gene promoters and silencing of repetitive elements at pericentromeric chromatin (Schotta et al., 2004; Wang et al., 2008), and a decrease in S5-phosphorylated RNAP II, which suggests lower transcriptional activity in *Mecp2*-null neurons. This condensation is likely mediated by the increase in linker histone H1 expression, which can compete with MeCP2 for chromatin binding and which doubles in expression in the absence of MeCP2 (Ghosh et al., 2010; Ishibashi et al., 2008; Skene et al., 2010). Notably, linker histone H1 also inhibits Dnmt3a-mediated *de novo* methylation at linker DNA (Takeshima et al., 2008) and exhibits higher affinity for linker DNA than MeCP2 by salt extraction (Nikitina et al., 2007b). The high levels of MeCP2 in the brain may thus displace transcriptionally repressive linker histone H1 bound to chromatin, and, through recruitment of interacting proteins, promote a more open chromatin environment at long neuronal genes that is not only transcriptionally efficient and leads to higher levels of gene expression, but may be more open to activity-dependent methylation changes that are important for learning, memory, and synaptic plasticity (Feng et al., 2010; Wu et al., 2010; Zovkic et al., 2013).

In a recent review of the RTT research field, Adrian Bird and colleagues proposed that "for a model of MeCP2 function to properly describe the pathology of RTT, it must explain how the same syndrome can arise from both null mutations and missense mutations predominantly in two discrete regions of the protein" (Lyst and Bird, 2015). Thus a mechanistic model of RTT must also be able to explain the severe phenotypic effects associated with mutations in the MBD of MeCP2, and the relatively milder phenotypic effects and onset of symptoms associated with mutations in the TRD of MeCP2. Using these criteria, the downregulation of long genes in neurons, which is dependent on differences in binding affinity using our T158M and R106W mice, may not be sufficient to *initiate* the onset of RTT symptoms. This is because missense mutations in the TRD of MeCP2, as well as C-terminal truncations, lead to a more gradual onset of RTT phenotypes over time without effecting changes in binding affinity to methylated DNA (Cuddapah et al., 2014). Given the correlations I established between transcriptional and phenotypic deficits,

the gradual onset of RTT phenotypes associated with these milder mutations must be accompanied by the gradual transcriptional downregulation of long genes. This is consistent with the progressive nature of RTT and the extensive increase in transcriptional deficits among neurons as T158M mice age over time. Thus a mechanistic model of RTT must also be consistent with the notion that long genes can be transcriptionally downregulated in the absence of mutations that disrupt MeCP2's affinity for methylated DNA. For this to occur, mutations that *do* affect MeCP2 binding to methylated DNA, chromatin, and potentially long genes (such as the T158M and R106W mutations) must increase the *rate* at which the symptoms progress, which is consistent with the reduced survival of R106W in comparison to T158M mice. Under this set of rules, the transcriptional efficiency of long neuronal genes is likely reduced in response to some other opposing instigating factor that initiates RTT phenotypes (discussed in more detail in section 4.4.2). MeCP2 binding to long neuronal genes may thus act to buffer that response, leading to an increase in transcriptional efficiency and proper neuronal function.

Although MeCP2 binding may be necessary to increase transcriptional efficiency, it may not be sufficient, and other functional aspect of MeCP2 may need to be considered in order to briefly speculate on a mechanism. The only phenotype that correlated with molecular severity between T158M mice and R106W mice was life span. As an aside, this suggests that the transcriptional efficiency of long neuronal genes may be particularly important for HoxB1+ neuronal function in the brainstem and spinal cord. This neuronal population helps mediate autonomic functions such as respiration and heart rate that are frequently affected in both RTT patients, and lifespan, which is severely affected in HoxB1+ conditional *Mecp2* mice (Ward et al., 2011). Most published *Mecp2* knockin mice have significantly longer life spans than *Mecp2*-null mice; the only two mouse models that closely resemble the *Mecp2*-null (11 weeks) and R106W (10 weeks) mice in reduced survival are R111G mice (11 weeks) and R270X mice (12 weeks) (Baker et al., 2013; Heckman et al., 2014). Of the two, R111G mutation resides in the MBD and severely disrupts binding to methylated DNA; hence the reduction in lifespan is not as surprising. The R270X is a C-terminal truncation that disrupts an AT-hook domain required to form higher

order oligomeric structures among nucleosomal arrays (Baker et al., 2013). In addition to a rapid onset of phenotypes that closely resemble *Mecp2*-null mice, this mutation retains MBD binding to methylated DNA *in vitro* and chromatin *in vivo*, exhibits no compensatory doubling of linker histone H1 expression, disrupts transcriptional repression activity by the TRD domain, and markedly reduces ATRX localization to pericentromeric chromatin in the nucleus (Baker et al., 2013).

Given a model where transcriptional downregulation of long genes correlates with phenotypes such as lifespan and is dependent on binding affinity, the R270X mutation suggests that the NCoR-SMRT interaction, and transcriptional repressor activity of MeCP2 in general, may be indispensible for promoting the expression of long neuronal genes. The ability to organize chromatin structure at long genes, perhaps through direct recruitment of the ATP-dependent chromatin remodeler and helicase ATRX, may be sufficient to increase transcriptional efficiency of long genes. Perhaps MeCP2 promotes de-condensation of chromatin at long neuronal genes in order to recruit ATRX, which subsequently maintains a more transcriptionally active environment though its nucleosome remodeling activity to increase the transcriptional efficiency of these genes. Such an environment may promote three-dimensional chromatin contacts that maintain the length-dependent expression of genes. Additional work should thus use HiC to investigate genome-wide three dimensional nuclear organization and chromatin structure in various *Mecp2*-mutant mouse models, which can now be performed at the single cell level and in intact nuclei for high resolution mapping (Nagano et al., 2013, 2015; Rao et al., 2014).

An A140V mouse model of x-linked mental retardation, which disrupts the ATRX interaction in mice without affecting MeCP2 binding or transcriptional repressor activity, has been shown to display reduced cell packing density and reduced dendritic complexity, similar to *Mecp2*-mutant mice (Jentarra et al., 2010). However, this mouse model also shows normal life span, weight gain, and other gross physiological features that distinguish it from *Mecp2*-mutant mice, suggesting that the molecular mechanisms between these two proteins are not clearly understood. Alternatively, if the transcriptional efficiency among long genes is reduced in the

A140V mouse model, perhaps the transcriptional downregulation of long genes alone may not be sufficient to mediate the gross clinical symptoms observed in RTT (discussed in more detail in section 4.4.2). Further work will be required to generate additional mutations in *Mecp2* for cell type-specific nuclear RNA sequencing and phenotypic assessment.

Together, these nuclear RNA sequencing experiments strongly implicate MeCP2 in facilitating the expression of long genes within adult neurons at the transcriptional level of gene regulation, with is consistent with prior studies using mouse brains as well as human-derived embryonic stem cells (Ben-Shachar et al., 2009; Chahrour et al., 2008; Li et al., 2013a). Although MeCP2 may exert effects akin to a transcriptional activator, further work will be required to determine the exact mechanism by which this occurs.

4.4 UPREGULATED GENES INITIATE RETT PHENOTYPES

4.4.1 Summary of Findings

Using aged-matched T158M and R106W mutant male mice at the onset of overt RTT phenotypes, I sequenced nuclear RNA from cortical excitatory and inhibitory neurons. Genome-wide, when MeCP2 function is generally impaired, the expression of short genes is predominantly upregulated in a length-dependent manner. This argues for MeCP2 in promoting the repression of short genes, regardless of cortical cell type. The degree to which short genes are upregulated *does not correlate* with the degree of MeCP2 impairment, in contrast to downregulated genes. Furthermore, genome-wide differences between these two mutations were not detected among genes that are transcriptionally engaged with RNA Polymerase II. Thus the active transcription of short genes in both excitatory and inhibitory neurons is indifferent, rather than sensitive, to the differences in MeCP2's ability to bind to chromatin *in vivo*; any change to MeCP2 binding or function may result in the upregulation of these genes. However, the degree to which genes are upregulated still worsens with age, which correlates with the progressive worsening of phenotypes when T158M mice are aged an additional 6 weeks.

The most striking feature observed among FDR-controlled upregulated genes was the similarity in molecular features between aged-matched T158M and R106W neurons. Upregulated DEGs in both cell types are typically low expressing, inhibitory-enriched genes (see section 4.2.1). However, the distributions of expression levels are similar between T158M and R106W mutant neurons, in contrast to the mutation-specific differences in expression levels among downregulated genes. This suggests that genes that are upregulated in the absence of MeCP2 function may lie upstream of the transcriptional consequences associated with RTT phenotypes. Furthermore, upregulated DEGs in both T158M and R106W excitatory neurons were similarly enriched for transcriptional regulators of gene expression. This association was also detected among cell autonomous DEGs in aged female T158M and R106W excitatory neurons, and was absent among non-cell autonomous DEGs. Taken together, this suggests that MeCP2 normally represses inhibitory-enriched transcriptional regulators, and that the cell autonomous de-repression of these genes occurs upstream of RTT-associated transcriptional effects. This is in agreement with genome-wide observations, where the increase in expression of inhibitory-enriched transcription factors, regardless of MeCP2 molecular severity, could promote the global expression of inhibitory-enriched genes, which are typically short in gene length (described in 4.2.1).

4.4.2 Broader Significance

Given the proposed roll for transcriptionally downregulated genes in modulating RTT phenotypes (see section 4.3.2) the ectopic expression of inhibitory-enriched transcription factors within excitatory neurons may serve the role of the “instigating factor” and have pronounced effects on the transcriptional efficiency of long neuronal genes *in vivo*. In excitatory neurons, these factors may act to negatively regulate the expression of long neuronal genes, such as by disrupting three-dimensional chromatin conformations that are mediated by precise combinations of transcription factors (Fullwood et al., 2009; Métivier et al., 2003). The strength with which these instigating factors downregulate long genes may rely on the accessibility of transcription factor sites at promoters or along the gene body of neuronal genes, which may be physically occluded

by MeCP2 occupancy itself, or by MeCP2 remodeling of nucleosome positioning with accessory factors (such as ATRX) (Kernohan et al., 2010, 2014; Thambirajah et al., 2012). Furthermore, as these instigating factors are continually expressed with time, the intranuclear protein concentration of these transcription factors will increase, which may lead to the formation of novel CTCF-mediated chromatin interactions across long neuronal gene bodies that can block RNA Polymerase II transcriptional elongation across a gene (Mayer et al., 2015). This likely explains the progressive but extensive downregulation of long neuronal genes that are observed among aged T158M mice in contrast to younger mice.

Through the association I established between cell type-enriched genes and gene lengths, these ectopically expressed inhibitory-enriched transcription factors are likely responsible for the genome-wide length-dependent increase in the transcription of short genes. These genes, which are similarly affected in both T158M and R106W mutant neurons regardless of GRO-seq gene filtering, most likely represent the transcriptional upregulation of inhibitory-enriched genes which tend to be shorter in gene length and would likely be responsive to the increased expression of their transcription factors. Given that inhibitory-enriched genes are associated with cellular respiration, metabolism, and mitochondrial function, the length-dependent transcription of these genes may be associated with mitochondrial abnormalities that have been observed in neurons of both post-mortem RTT patients and *Mecp2* mutant mice (Belichenko et al., 2009; Cornford et al., 1994; Großer et al., 2012), and would be consistent with reports in recent years that highlight the similarities in clinical features between RTT and metabolic or mitochondrial disorders (Buchovecky et al., 2013; De Filippis et al., 2015; Filosa et al., 2015; Müller and Can, 2014; Valenti et al., 2014).

Given my genome-wide observations, the same transcriptional mechanisms described above may also apply to inhibitory neurons. However, because these inhibitory-enriched transcription factors are being de-repressed in an inhibitory cell type, this might only result in strengthening pre-existing chromatin conformations, leading to an increase in the already high expression of inhibitory expressed genes. This likely explains why the majority of DEGs in 6 week

T158M and R106W interneurons are upregulated, in contrast to excitatory neurons where genes are predominantly downregulated. The transcriptional effects on long genes may not be as robust in interneurons because this cell type does not exhibit the same length-dependent transcriptional dynamics as excitatory neurons do for long neuronal genes.

The low expression of these instigating transcription factors suggest that these genes are infrequently transcribed, which supports a mechanism for these genes being de-repressed in the absence of MeCP2 function. Not surprisingly, this would require MeCP2's interactions with NCoR-SMRT and Sin3a histone deacetylase co-repressor complexes. However, models of co-repressor function are continually evolving. Rather than contributing to long-term gene repression, which is likely maintained by histone post-translational modifications such as H3K27me3 and H3K9me3, co-repressor activity can be a dynamically regulated process to quickly modulate the expression of genes and the chromatin environment (Perissi et al., 2010). This is consistent with genome-wide studies of histone deacetylase (HDAC) and histone acetyltransferase (HAT) occupancy in human cells, which found that the majority of HDACS in the human genome function to reset chromatin by removing acetylation at active genes, including genes that may be primed for activation due to the presence of H3K4me3 (Perissi et al., 2010; Wang et al., 2009).

One of the mechanisms by which this can occur is through the cyclic recruitment of co-repressor complexes to genes. Cyclic recruitment to lowly expressed genes may result in a gene promoter transitioning between an active acetylated "expression" state, and an inactive deacetylated "repressed" state (Raj and van Oudenaarden, 2008). This can lead to stochastic transcriptional "bursts" of gene expression whose frequency can be regulated by signaling pathways that produce biologically-relevant outcomes (Raj and van Oudenaarden, 2008; Raj et al., 2006) and can occur in mammalian tissues *in vivo* (Bahar Halpern et al., 2015; Featherstone et al., 2016). In neurons, stochastic regulation of gene expression may be beneficial for modulating the activity-dependent expression of ion channels that affect the intrinsic excitability of pyramidal neurons, which are known to vary in electrophysiological properties from cell to cell

(Fuzik et al., 2016). In support of this notion, biophysical measurements of networked mitrial cells in the mouse olfactory bulb reveal that increases in the heterogeneity of intrinsic firing properties between individual cells were associated with a two-fold increase in the amount of information that could be processed by these neuronal networks (Padmanabhan and Urban, 2010).

Thus, a model supporting the ectopic upregulation of inhibitory-enriched genes in the absence of MeCP2 function may also support a role for MeCP2 in regulating stochastic gene expression, which can effect changes that affect overall neuronal network activity (Blackman et al., 2012; Maffei and Fontanini, 2009; Nguyen et al., 2012; Qiu et al., 2012; Zhong et al., 2012). Such a model incorporates a controlled de-repression of inhibitory-enriched transcription factors that act to negatively regulate long neuronal genes in response to excessive neuronal activity. Notably, MeCP2 is phosphorylated at S421 and S424 and dephosphorylated at S80 in an activity-dependent manner, which are thought to alter its affinity for methylated DNA and chromatin *in vivo* (Cohen et al., 2011; Tao et al., 2009; Zhou et al., 2006). Under this cyclic repression model, subtle changes in MeCP2's binding affinity, such as those mediated by phosphorylating and dephosphorylating events, would increase its rate of dissociation from chromatin. I expect the effect would be strongest among genes with low densities of methyl-CpG or methyl-CpA, in contrast to long neuronal genes with have high densities of methyl-CpG or high methyl-CpA. This model predicts that physiological activity-dependent changes in MeCP2 binding affinity for methylated DNA are subtle, as T158M and R106W mutations, which partially or fully reduce MeCP2's affinity for methylated DNA, show the same genome-wide upregulation of genes despite their strong differences in molecular severity.

Together, these nuclear RNA sequencing experiments indicate that MeCP2 mutations induce the ectopic expression transcriptional regulators within adult neurons that may act to initiate RTT symptoms. This requires the transcriptional repressor activity of MeCP2, which is well established in the RTT field and can be regulated in an activity dependent manner (Ebert et al., 2013; Lyst et al., 2013; Nan et al., 1997, 1998; Tao et al., 2009). Alterations to MeCP2 binding, or to its ability to interact and recruit HDAC-containing co-repressor complexes, may lead to the

same upregulation of these transcriptional regulators. Therefore, the increased expression of these transcription factors is also consistent with the mutational spectrum of RTT. Interestingly, this likely explains why the aforementioned A140V mouse model does not display extensive RTT phenotypes. The A140V mutation, which has been directly compared the R106W mutation in heterologous cells, does not alter MeCP2 binding, nor does it affect MeCP2-mediated transcriptional repression (Kudo et al., 2002; Nan et al., 2007). This provides strong genetic evidence supporting the upregulation of these transcriptional regulators as instigators of the RTT phenotype. Further work will be necessary to determine if the pathological upregulation of these instigators is part of a much larger physiological role for MeCP2 in mediating stochastic gene expression to effect changes in synaptic plasticity.

4.5 POST-TRANSCRIPTIONAL COMPENSATION OF RETT PHENOTYPES

4.5.1 Summary of Findings

In an effort to repeat the findings of several recent reports detailing length-dependent repression of long genes by MeCP2 (Brown et al., 2015; Gabel et al., 2015; Sugino et al., 2014), I performed a genome-wide analysis assessing the mean $\log_2$ fold change among T158M and R106W mice using both nascent RNA (chromatin-associated), nuclear RNA (chromatin-associated + nucleoplasm), and whole cell RNA (chromatin-associated + nucleoplasm + cytoplasm). I found that, though whole cell RNA recapitulated the length-dependent increase in the expression of long genes that was described in the aforementioned studies, nascent and nuclear RNA fractions show the opposite trend. In the nucleus, long genes are predominantly downregulated, whereas short genes are predominantly upregulated. These results show that MeCP2-mediated gene expression changes in the nucleus reflect true transcriptional changes by MeCP2, whereas MeCP2-mediated gene expression changes using whole cell RNA reflect products of post-transcriptional compensation.

4.5.2 Broader Significance

Multiple studies have corroborated the extensive differences in transcriptome profiles between subcellular compartments (Bhatt et al., 2012; Djebali et al., 2012; Trask et al., 2009; Werner and Ruthenburg, 2015). However, the extensive length-dependent differences in gene expression between nascent, nuclear, and whole cell RNA fractions indicate that MeCP2 mediates biologically relevant transcriptional changes on a genome-wide scale that the cell actively attempts to minimize. Although alterations to nuclear export may account for some of the differences in fold change between subcellular compartments, I think a more likely mechanism lies among ribonucleoprotein (RNP) granules in the cytoplasm. Neurons make extensive use of post-transcriptional regulatory processes to mediate activity-dependent translation at dendrites, synaptic terminals, and other areas of the cell (Buxbaum et al., 2015). RNP granules are non-membranous organelle structures that serve key roles in mRNA homeostasis (Li et al., 2013b), participating in the localization, modification, and decay of mRNAs, especially during times of cellular stress. RNP granules include P-bodies, which represent sites of mRNA degradation, translational repression, non-translating mRNAs and RNA-binding proteins (Jain and Parker, 2013). Another class of RNP granule includes stress granules, which are sites where translationally stalled or inert RNAs can be stored in the cytoplasm until needed, especially during translational perturbations (Anderson and Kedersha, 2008).

Given these structures, it's likely that short genes are exhibiting higher mRNA turnover and being degraded more frequently miRNA-mediated mechanisms that ultimately target these short RNAs to P-bodies for degradation. Likewise, stress granules may accumulate in the cytoplasm to reduce mRNA turnover and mediate the cytoplasmic retention of long synaptic genes as a way of conserving a limited resource to maintain overall neuronal function. The incorporation of long genes into stress granules may be of particular relevance to RTT, which exhibit severe translational deficiencies that are associated with synaptic deficits in plasticity and are thought to underlie the small soma size of neurons; likewise, *Mecp2*-null ESC-derived neurons also display reduced ribonucleotide incorporation of ribosomal RNAs in the nucleus,

which would lead to an overall cellular decrease in ribosomal content (Li et al., 2013a; Yazdani et al., 2012).

However, it remains unknown whether post-transcriptional compensation is truly beneficial, or whether it is effectual in minimizing RTT phenotypes. It would also be interesting to know how this process is initiated and regulated, particularly during the regression phase of RTT and onset of symptoms. Of extreme interest, supplementary figure 2a and 2b of (Gabel et al., 2015) performed genome-wide analysis on *Mecp2*-null, R270X, and G273X mouse models at 4 week and 9 weeks of age using whole cell RNA. Interestingly, although *Mecp2*-null mice show a mean decrease in the expression of short genes at both 4 weeks and 9 weeks of age, the R270X and G273X mouse models show a mean increase in the expression of short genes at 4 weeks of age, which are then downregulated (compensated) by 9 weeks of age. This suggests that post-transcriptional compensation in the cytoplasm initiates during the onset of RTT phenotypes.

It also remains an open question whether post-transcriptionally compensated RNAs in *Mecp2*-mutant neurons are translated at levels similar to wild-type neurons, or whether these long neuronal RNAs are sequestered but remain quiescent given the cell autonomous deficiency of ribosomal RNAs and translational activity. More work is needed to understand the dynamics of translational regulation in *Mecp2*-mutant neurons.

4.6 IT'S ALL ABOUT CONTEXT: A MODEL FOR MECP2 IN RETT SYNDROME

4.6.1 The Model

Within the context of an excitatory pyramidal neuron, and at basal conditions, MeCP2 remains stably bound to long excitatory-enriched neuronal genes with high densities of methyl-CpG/CpA (modulator genes) to promote efficient and stable expression, which may occur through increasing nucleosome mobility along gene bodies or participation in three dimensional chromatin structures. At the same time, MeCP2 dynamically binds to genes encoding inhibitory-enriched

transcriptional regulators with low densities of methyl-CpG/CpA (instigator genes) in order to mediate their stochastic expression.

In response to normal neuronal activity, MeCP2 is post-translationally phosphorylated to subtly reduce its binding affinity for methylated DNA. This decreases its residence time with chromatin, leading to the stochastic upregulation and expression of instigators, which then act to negatively regulate the expression of modulator genes. However, MeCP2 remains largely bound at modulator genes because of higher densities of methyl-CpG/CpA and cooperative binding with other surrounding MeCP2 molecules that promote higher-order structures. Given that the distribution of methyl-CpG does not strongly correlate with expression level (Lister et al., 2013; Mo et al., 2015), modulator genes with high levels of CpG-bound MeCP2 and high transcriptional activity (Group A modulators) are easily able to buffer the negative effects of the instigators. However, because methyl-CpA is positively correlated with gene repression (Mo et al., 2015), the presence of MeCP2 may keep these types of modulator genes from being fully repressed and results in relatively low levels of transcriptional activity (Group B modulators). The paucity of high transcriptional activity among Group B modulator genes makes these genes particularly susceptible to stochastic influences of instigators, resulting in the stochastic downregulation of Group B modulators. After a certain amount of time, the system resets, de-phosphorylated MeCP2 resumes its normal dynamic binding at instigator genes, and Group B modulator genes either return to baseline, or are epigenetically modified to retain the new expression state.

In response to excessive amounts of neuronal activity, MeCP2 is post-translationally phosphorylated genome-wide, resulting in global decreases in binding affinity. The proportion of MeCP2 that gets phosphorylated may increase given how long neuronal activity persists, and the global reduction in binding affinity may be enough to overcome the influence of high density methyl-CpG/CpA and cooperative binding. Again, this results in the stochastic upregulation and expression of instigators, which then act to negatively regulate the expression of modulator genes. However, Group A and Group B modulators are now partially or fully depleted of MeCP2,

depending on the degree of neuronal activity, and are thus fully susceptible to the negative regulatory effects of instigators. This leads to a global downregulation of neuronal and synaptic genes related to neuronal plasticity in order to scale back the excessive excitatory input that the neuron is receiving. After neuronal activity has passed, the system resets, de-phosphorylated MeCP2 resumes its normal dynamic binding at instigator genes, and Group A and B modulator genes either to baseline, or are epigenetically modified to retain the new expression state.

In a T158M RTT neuron, MeCP2's residence time with chromatin is constitutively reduced regardless of neuronal activity, leading to the constitutive upregulation and expression of instigators, which then act to negatively regulate the expression of modulator genes. However, although MeCP2 T158M exhibits significantly reduced binding to methylated DNA, it still has a high enough residence with chromatin to exhibit partial binding to Group A and Group B modulators. The expression of Group B modulators is consistently downregulated because of their increased sensitivity to the negative influences of instigators. Group A modulators, exhibit a partial buffering capacity against instigators due to partial binding by MeCP2 T158M, but are still more resistant to instigators because of their higher levels of transcriptional activity. However, as the amount of instigator protein increases over time, Group A modulators gradually decrease in transcriptional activity, with lower expressing genes being more susceptible to downregulation than higher expressing genes. This results in a gradual downregulation of long neuronal genes over time.

In a R106W RTT neuron, MeCP2's residence time with chromatin is constitutively reduced regardless of neuronal activity, leading to the constitutive upregulation and expression of instigators, which then act to negatively regulate the expression of modulator genes. However, MeCP2 R106W exhibits no affinity for methylated DNA and minimal residence with chromatin. The transcriptional activity of Group B modulators is consistently downregulated because of their increased sensitivity to the negative influences of instigators. Group A modulators have no buffering capacity against instigators aside from higher levels of transcriptional activity. However,

as the amount of instigator protein increases over time, Group A modulators more rapidly decrease in transcriptional activity, with lower expressing genes being more susceptible to downregulation than higher expressing genes. This results in a faster rate of downregulation of long neuronal genes over time, especially when compared to T158M RTT neurons. The rapid downregulation or change in network activity may induce the expression of immediate early genes and secretory molecules in an effort to compensate and stimulate synaptic plasticity processes.

4.6.2 Implications

This model has potential implications for how neurons may respond to neuronal activity at the transcriptional level. Given that mRNA abundance does not strongly correlate with protein abundance, Group B modulators may encode genes that are very efficient at being translated into protein, which can occur with stochastically expressed genes (Raj and van Oudenaarden, 2008). Thus these genes may encode proteins whose expression at the plasma membrane can be dynamically regulated at the transcriptional level. In the context of the datasets generated in this study, Group B modulators represent the R106W downregulated DEGs, which were significantly enriched for ion channels and transporter activity. Transcriptional changes in genes encoding ion channels are known to occur in response to neuronal activity; furthermore, neurons use activity-dependent feedback mechanisms to regulate receptors and overall membrane conductances (O'Leary et al., 2014). Computational models show that the most efficient way for homeostatic mechanisms to occur is to have a "master regulator" that can both detect changes in neuronal activity and alter neuronal conductance in response (O'Leary et al., 2014). One type of homeostatic feedback mechanism is called synaptic scaling, an activity-dependent and transcription-mediated process that scales the expression of ion channels to change the neuron's long-term intrinsic excitability until a sense of balance is reached (Turrigiano, 2011). Interestingly, MeCP2 has been implicated in synaptic scaling, as MeCP2-deficient neurons demonstrate a failure to scale their neuronal conductance in line with excessive or diminished neuronal activity (Blackman et al., 2012; Qiu et al., 2012; Zhong et al., 2012). Furthermore, this requires MeCP2

phosphorylation on S421 and S424, and thus requires a reduction in MeCP2 affinity for chromatin *in vivo* (Zhong et al., 2012).

Therefore, the model that I propose to explain the observations I see within T158M and R106W mice also provides a novel transcriptional mechanism for how MeCP2-mediated synaptic scaling could occur through activation of a negative feedback circuit in response to neuronal activity.

4.6.3 Therapeutic Significance

The existence of post-transcriptionally compensated whole cell RNA, which is enriched for cytoplasmic polyadenylated and ribosome-associated RNAs and is thus used in nearly every transcriptional study involving MeCP2 (Ben-Shachar et al., 2009; Chahrour et al., 2008; Colantuoni et al., 2001; Jordan et al., 2007; Kriaucionis et al., 2006; Li et al., 2013a; Mellén et al., 2012; Nuber et al., 2005; Sugino et al., 2014; Tudor et al., 2002; Urdinguio et al., 2008), may force a re-assessment as to how MeCP2 should be studied in the future, and does caution the practice of inferring protein function from changes in gene expression. For example, the existence of post-transcriptional mechanisms to minimize MeCP2-mediated changes may explain why some RTT patients, after experience the regressive phase of the syndrome, enter a pseudo-stationary phase where they partially recover lost developmental skills, and may underlie the slower rate of symptom onset in female, rather than male mice. However, by assuming from whole cell preparations that MeCP2 mediates transcriptional repression, advocating the use of topoisomerase inhibitors to treat RTT patients may do more potential harm than good.

4.7 LIMITATIONS ASSOCIATED WITH WORK

One potential limitation with this work includes the phenotypic characterization of the T158M and R106W mice. Although these mice exhibit extensive differences in molecular and transcriptional phenotypes, the physiological and behavioral characterization of these mice is limited and lacking. More robust behavioral characterizations are required to identify additional physiological and behavioral phenotypes that correlate with transcriptional changes. This is

especially important for female mice, which are now ideal for investigating how behavioral and phenotypic changes correlate with changes in X-inactivation ratios of wild-type and mutant neurons, especially when R106W mice are more prone to non-cell autonomous changes than T158M mice. This may unveil new molecular mechanisms that correlate with X-inactivation and the non-cell autonomous promotion of cellular health, using genetics to lead to the identification of novel secreted molecules for future therapeutic use.

Another limitation involves the cortical cell types that were profiled. Although two transcriptionally distinct cell types were isolated, “excitatory” and “inhibitory” neurons are arbitrary classification schemes in the broadest sense. These two cell types encompass a rich and extraordinarily wide variety of cell types that are only now coming to light with the advent of new technology, including single cell sequencing (Fuzik et al., 2016; Tasic et al., 2016; Usoskin et al., 2015). Each individual neuron is its own cell type, reflecting a unique electrophysiological signature and gene expression profile that is functionally distinct from another neuron of the same “cell type” (Fuzik et al., 2016). Thus many of the cell type-specific associations I discovered may only represent transcriptional dynamics among a functionally heterogeneous group of cells at a population level. Future studies will be needed to determine how these dynamics are reflected at the single cell level, which also has the advantage of expanding the number of cell types being profiled at any one time for more accurate comparisons and biological insights. Single cell sequencing may be especially important for MeCP2, given the large amount of variability in transcriptional activity that is observed at the population level.

Another limitation is the lack of MeCP2 binding sites to confirm whether many of the hypotheses I generated in this dissertation holds true. This could not be helped given the extensive amount of time I spent over the years trying to optimize cell type-specific MeCP2 ChiP-seq. This has been a major limitation for the RTT field, given the global distribution of MeCP2. However, I believe that the difficulty results from the biology of the protein itself, and limitations on how ChIP is traditionally performed. The model that I propose in this dissertation predicts that, although MeCP2 may be found throughout the genome, its binding may be highly dynamic, and

its resident interactions with chromatin may differ depending on gene length, methylation density, and cell type in order to dynamically modulate gene expression in response to neuronal activity. Therefore, traditional methods of ChIP may not be sufficient to capture this dynamic binding of MeCP2. Novel strategies that assess binding dynamics of proteins over time, such as competitive ChIP, may be the path going forward to test how differences in MeCP2 binding affinity and dynamics can influence gene expression in an activity-dependent manner, or in response to RTT-associated mutations.

BIBLIOGRAPHY

Abuhatzira, L., Makedonski, K., Kaufman, Y., Razin, A., and Shemer, R. (2007). MeCP2 deficiency in the brain decreases BDNF levels by REST/CoREST-mediated repression and increases TRKB production. Epigenetics *2*, 214–222.

Adachi, M., Autry, A.E., Covington, H.E., and Monteggia, L.M. (2009). MeCP2-Mediated Transcription Repression in the Basolateral Amygdala May Underlie Heightened Anxiety in a Mouse Model of Rett Syndrome. J. Neurosci. *29*, 4218–4227.

Adams, V.H., McBryant, S.J., Wade, P.A., Woodcock, C.L., and Hansen, J.C. (2007). Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. J. Biol. Chem. *282*, 15057–15064.

Agarwal, N., Becker, A., Jost, K.L., Haase, S., Thakur, B.K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H., and Cardoso, M.C. (2011). MeCP2 Rett mutations affect large scale chromatin organization. Hum. Mol. Genet. *20*, 4187–4195.

Alvarez-Saavedra, M., Sáez, M.A., Kang, D., Zoghbi, H.Y., and Young, J.I. (2007). Cell-specific expression of wild-type MeCP2 in mouse models of Rett syndrome yields insight about pathogenesis. Hum. Mol. Genet. *16*, 2315–2325.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.

Anderson, P., and Kedersha, N. (2008). Stress granules: the Tao of RNA triage. Trends Biochem. Sci. *33*, 141–150.

Archer, H., Evans, J., Leonard, H., Colvin, L., Ravine, D., Christodoulou, J., Williamson, S., Charman, T., Bailey, M.E.S., Sampson, J., et al. (2007). Correlation between clinical severity in patients with Rett syndrome with a p.R168X or p.T158M MECP2 mutation, and the direction and degree of skewing of X-chromosome inactivation. J. Med. Genet. *44*, 148–152.

Ariani, F., Hayek, G., Rondinella, D., Artuso, R., Mencarelli, M.A., Spanhol-Rosseto, A., Pollazzon, M., Buoni, S., Spiga, O., Ricciardi, S., et al. (2008). FOXG1 Is Responsible for the Congenital Variant of Rett Syndrome. Am. J. Hum. Genet. *83*, 89–93.

Armstrong, D.D. (2005). Neuropathology of Rett Syndrome. J. Child Neurol. *20*, 747–753.

Artuso, R., Mencarelli, M.A., Polli, R., Sartori, S., Ariani, F., Pollazzon, M., Marozza, A., Cilio, M.R., Specchio, N., Vigevano, F., et al. (2010). Early-onset seizure variant of Rett syndrome: definition of the clinical diagnostic criteria. Brain Dev. *32*, 17–24.

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty Gene Expression in the Intact Mammalian Liver. Mol. Cell *58*, 147–156.

Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., and Zoghbi, H.Y. (2013). An AT-Hook Domain in MeCP2 Determines the Clinical Course of Rett Syndrome and Related Disorders. Cell *152*, 984–996.

Ballestar, E., Yusufzai, T.M., and Wolffe, A.P. (2000). Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. Biochemistry (Mosc.) *39*, 7100–7106.

Balmer, D., Goldstine, J., Rao, Y.M., and LaSalle, J.M. (2002). Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. J. Mol. Med. *81*, 61–68.

Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. Cell *153*, 480–492.

Beaulieu, C. (1993). Numerical data on neocortical neurons in adult rat, with special reference to the GABA population. Brain Res. *609*, 284–292.

Bebbington, A., Anderson, A., Ravine, D., Fyfe, S., Pineda, M., de Klerk, N., Ben-Zeev, B., Yatawara, N., Percy, A., Kaufmann, W.E., et al. (2008). Investigating genotype-phenotype relationships in Rett syndrome using an international data set. Neurology *70*, 868–875.

Becker, A., Zhang, P., Allmann, L., Meilinger, D., Bertulat, B., Eck, D., Hofstaetter, M., Bartolomei, G., Hottiger, M.O., Schreiber, V., et al. (2016). Poly(ADP-ribosyl)ation of Methyl CpG Binding Domain Protein 2 Regulates Chromatin Structure. J. Biol. Chem. *291*, 4873–4881.

Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., and Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc. Natl. Acad. Sci. U. S. A. *95*, 14173–14178.

Belichenko, N.P., Belichenko, P.V., Li, H.H., Mobley, W.C., and Francke, U. (2008). Comparative study of brain morphology in Mecp2 mutant mouse models of Rett syndrome. J. Comp. Neurol. *508*, 184–195.

Belichenko, P.V., Oldfors, A., Hagberg, B., and Dahlström, A. (1994). Rett syndrome: 3-D confocal microscopy of cortical pyramidal dendrites and afferents. Neuroreport *5*, 1509–1513.

Belichenko, P.V., Wright, E.E., Belichenko, N.P., Masliah, E., Li, H.H., Mobley, W.C., and Francke, U. (2009). Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of rett syndrome: Evidence for disruption of neuronal networks. J. Comp. Neurol. *514*, 240–258.

Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. Genet. *12*, 554–564.

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum. Mol. Genet. *18*, 2431–2442.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. Cell *150*, 279–290.

Blackman, M.P., Djukic, B., Nelson, S.B., and Turrigiano, G.G. (2012). A Critical and Cell-Autonomous Role for MeCP2 in Synaptic Scaling Up. J. Neurosci. *32*, 13529–13536.

Bloodgood, B.L., Sharma, N., Browne, H.A., Trepman, A.Z., and Greenberg, M.E. (2013). The activity-dependent transcription factor NPAS4 regulates domain-specific inhibition. Nature *503*, 121–125.

Boer, E. de, Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci USA *100*, 7480–5.

Brown, K., Selfridge, J., Lagger, S., Connelly, J., Sousa, D.D., Kerr, A., Webb, S., Guy, J., Merusi, C., Koerner, M.V., et al. (2015). The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. Hum. Mol. Genet. ddv496.

Buchovecky, C.M., Turley, S.D., Brown, H.M., Kyle, S.M., McDonald, J.G., Liu, B., Pieper, A.A., Huang, W., Katz, D.M., Russell, D.W., et al. (2013). A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. Nat. Genet. *45*, 1013–1020.

Bustin, M., Catez, F., and Lim, J.-H. (2005). The Dynamics of Histone H1 Function in Chromatin. Mol. Cell *17*, 617–620.

Buxbaum, A.R., Yoon, Y.J., Singer, R.H., and Park, H.Y. (2015). Single-molecule insights into mRNA dynamics in neurons. Trends Cell Biol. *25*, 468–475.

Cartron, P.-F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. Genes Cancer *4*, 235–241.

Castro, J., Garcia, R.I., Kwok, S., Banerjee, A., Petravicz, J., Woodson, J., Mellios, N., Tropea, D., and Sur, M. (2014). Functional recovery with recombinant human IGF1 treatment in a mouse model of Rett Syndrome. Proc. Natl. Acad. Sci. U. S. A. *111*, 9941–9946.

Chahrour, M., and Zoghbi, H.Y. (2007). The story of Rett syndrome: from clinic to neurobiology. Neuron *56*, 422–37.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. Science *320*, 1224–1229.

Chang, Q., Khare, G., Dani, V., Nelson, S., and Jaenisch, R. (2006). The Disease Progression of Mecp2 Mutant Mice Is Affected by the Level of BDNF Expression. Neuron *49*, 341–348.

Chao, H.-T., Zoghbi, H.Y., and Rosenmund, C. (2007). MeCP2 Controls Excitatory Synaptic Strength by Regulating Glutamatergic Synapse Number. Neuron *56*, 58–65.

Chao, H.-T., Chen, H., Samaco, R.C., Xue, M., Chahrour, M., Yoo, J., Neul, J.L., Gong, S., Lu, H.-C., Heintz, N., et al. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. Nature *468*, 263–269.

Chapleau, C.A., Calfa, G.D., Lane, M.C., Albertson, A.J., Larimore, J.L., Kudo, S., Armstrong, D.L., Percy, A.K., and Pozzo-Miller, L. (2009). Dendritic spine pathologies in hippocampal

pyramidal neurons from Rett syndrome brain and after expression of Rett-associated MECP2 mutations. Neurobiol. Dis. *35*, 219–233.

Charman, T., Neilson, T.C.S., Mash, V., Archer, H., Gardiner, M.T., Knudsen, G.P.S., McDonnell, A., Perry, J., Whatley, S.D., Bunyan, D.J., et al. (2005). Dimensional phenotypic analysis and functional categorisation of mutations reveal novel genotype–phenotype associations in Rett syndrome. Eur. J. Hum. Genet. *13*, 1121–1130.

Chatr-Aryamontri, A., Breitkreutz, B.-J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., et al. (2013). The BioGRID interaction database: 2013 update. Nucleic Acids Res. *41*, D816-823.

Chen, L., Chen, K., Lavery, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc. Natl. Acad. Sci. *112*, 5509–5514.

Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet *27*, 327–31.

Chen, S.H., Chan, N.-L., and Hsieh, T. (2013). New Mechanistic and Functional Insights into DNA Topoisomerases. Annu. Rev. Biochem. *82*, 139–170.

Chen, W.G., Chang, Q., Lin, Y., Meissner, A., West, A.E., Griffith, E.C., Jaenisch, R., and Greenberg, M.E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science *302*, 885–889.

Chen, Y., Shin, B.-C., Thamotharan, S., and Devaskar, S.U. (2014). Differential methylation of the micro-RNA 7b gene targets postnatal maturation of murine neuronal Mecp2 gene expression. Dev. Neurobiol. *74*, 407–425.

Cheval, H., Guy, J., Merusi, C., Sousa, D.D., Selfridge, J., and Bird, A. (2012). Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. Hum. Mol. Genet. *21*, 3806–3814.

Christodoulou, J., Grimm, A., Maher, T., and Bennetts, B. (2003). RettBASE: The IRSA MECP2 variation database-a new mutation database in evolution. Hum. Mutat. *21*, 466–472.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annu. Rev. Biochem. *78*, 273–304.

Cohen, S., Gabel, H.W., Hemberg, M., Hutchinson, A.N., Sadacca, L.A., Ebert, D.H., Harmin, D.A., Greenberg, R.S., Verdine, V.K., Zhou, Z., et al. (2011). Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. Neuron *72*, 72–85.

Colantuoni, C., Jeon, O.-H., Hyder, K., Chenchik, A., Khimani, A.H., Narayanan, V., Hoffman, E.P., Kaufmann, W.E., Naidu, S., and Pevsner, J. (2001). Gene Expression Profiling in Postmortem Rett Syndrome Brain: Differential Gene Expression and Patient Classification. Neurobiol. Dis. *8*, 847–865.

Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., Sweatt, J.D., and Zoghbi, H.Y. (2004). Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. Hum. Mol. Genet. *13*, 2679–2689.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845–1848.

Cornford, M.E., Philippart, M., Jacobs, B., Scheibel, A.B., and Vinters, H.V. (1994). Neuropathology of Rett syndrome: case report with neuronal and mitochondrial abnormalities in the brain. J. Child Neurol. *9*, 424–431.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*, 6404.

Coy, J.F., Sedlacek, Z., Bächner, D., Delius, H., and Poustka, A. (1999). A Complex Pattern of Evolutionary Conservation and Alternative Polyadenylation Within the Long 3′-Untranslated Region of the Methyl-CpG-Binding Protein 2 Gene (MeCP2) Suggests a Regulatory Role in Gene Expression. Hum. Mol. Genet. *8*, 1253–1262.

Cuddapah, V.A., Pillai, R.B., Shekar, K.V., Lane, J.B., Motil, K.J., Skinner, S.A., Tarquinio, D.C., Glaze, D.G., McGwin, G., Kaufmann, W.E., et al. (2014). Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. J. Med. Genet. *51*, 152–158.

Dani, V.S., Chang, Q., Maffei, A., Turrigiano, G.G., Jaenisch, R., and Nelson, S.B. (2005). Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett Syndrome. Proc. Natl. Acad. Sci. U. S. A. *102*, 12560–12565.

De Filippis, B., Valenti, D., Chiodi, V., Ferrante, A., de Bari, L., Fiorentini, C., Domenici, M.R., Ricceri, L., Vacca, R.A., Fabbri, A., et al. (2015). Modulation of Rho GTPases rescues brain mitochondrial dysfunction, cognitive deficits and aberrant synaptic plasticity in female mice modeling Rett syndrome. Eur. Neuropsychopharmacol. *25*, 889–901.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. Nature *485*, 376–380.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dragich, J.M., Kim, Y.-H., Arnold, A.P., and Schanen, N.C. (2007). Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. J. Comp. Neurol. *501*, 526–542.

Driegen, S., Ferreira, R., Zon, A. van, Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res *14*, 477–82.

Ebert, D.H., and Greenberg, M.E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. Nature *493*, 327–337.

Ebert, D.H., Gabel, H.W., Robinson, N.D., Kastan, N.R., Hu, L.S., Cohen, S., Navarro, A.J., Lyst, M.J., Ekiert, R., Bird, A.P., et al. (2013). Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. Nature *499*, 341–345.

Emes, R.D., Pocklington, A.J., Anderson, C.N.G., Bayes, A., Collins, M.O., Vickers, C.A., Croning, M.D.R., Malik, B.R., Choudhary, J.S., Armstrong, J.D., et al. (2008). Evolutionary expansion and anatomical specialization of synapse proteome complexity. Nat. Neurosci. *11*, 799–806.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199–1212.

Featherstone, K., Hey, K., Momiji, H., McNamara, A.V., Patist, A.L., Woodburn, J., Spiller, D.G., Christian, H.C., McNeilly, A.S., Mullins, J.J., et al. (2016). Spatially coordinated dynamic gene transcription in living pituitary tissue. eLife *5*, e08494.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat. Neurosci. *13*, 423–430.

Filosa, S., Pecorelli, A., D'Esposito, M., Valacchi, G., and Hajek, J. (2015). Exploring the possible link between MeCP2 and oxidative stress in Rett syndrome. Free Radic. Biol. Med. *88*, *Part A*, 81–90.

Free, A., Wakefield, R.I.D., Smith, B.O., Dryden, D.T.F., Barlow, P.N., and Bird, A.P. (2001). DNA Recognition by the Methyl-CpG Binding Domain of MeCP2. J. Biol. Chem. *276*, 3353–3360.

Friez, M.J., Jones, J.R., Clarkson, K., Lubs, H., Abuelo, D., Bier, J.-A.B., Pai, S., Simensen, R., Williams, C., Giampietro, P.F., et al. (2006). Recurrent Infections, Hypotonia, and Mental Retardation Caused by Duplication of MECP2 and Adjacent Region in Xq28. Pediatrics *118*, e1687–e1695.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Furey, T.S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nat. Rev. Genet. *13*, 840–852.

Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S., and Harkany, T. (2016). Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. Nat. Biotechnol. *34*, 175–183.

Fyffe, S.L., Neul, J.L., Samaco, R.C., Chao, H.-T., Ben-Shachar, S., Moretti, P., McGill, B.E., Goulding, E.H., Sullivan, E., Tecott, L.H., et al. (2008). Deletion of Mecp2 in Sim1-expressing neurons reveals a critical role for MeCP2 in feeding behavior, aggression, and the response to stress. Neuron *59*, 947–958.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature *522*, 89–93.

del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M., et al. (2006). Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. Genet. Med. Off. J. Am. Coll. Med. Genet. *8*, 784–792.

Gemelli, T., Berton, O., Nelson, E.D., Perrotti, L.I., Jaenisch, R., and Monteggia, L.M. (2006). Postnatal loss of methyl-CpG binding protein 2 in the forebrain is sufficient to mediate behavioral aspects of Rett syndrome in mice. Biol. Psychiatry *59*, 468–476.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. J. Biol. Chem. *278*, 32181–32188.

Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Gierasch, L.M., and Woodcock, C.L. (2008). Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. J Biol Chem *283*, 20523–34.

Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Shlyakhtenko, L.S., and Woodcock, C.L. (2010). MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Mol. Cell. Biol. *30*, 4656–4670.

Giacometti, E., Luikenhuis, S., Beard, C., and Jaenisch, R. (2007). Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. Proc. Natl. Acad. Sci. *104*, 1931–1936.

Gianakopoulos, P.J., Zhang, Y., Pencea, N., Orlic-Milacic, M., Mittal, K., Windpassinger, C., White, S.-J., Kroisel, P.M., Chow, E.W.C., Saunders, C.J., et al. (2012). Mutations in MECP2 exon 1 in classical Rett patients disrupt MECP2_e1 transcription, but not transcription of MECP2_e2. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *159B*, 210–216.

Goebbels, S., Bormuth, I., Bode, U., Hermanson, O., Schwab, M.H., and Nave, K.-A. (2006). Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. Genes. N. Y. N 2000 *44*, 611–621.

Goffin, D., and Zhou, Z. (Joe) (2012). The neural circuit basis of Rett syndrome. Front. Biol. *7*, 428–435.

Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.-T.J., Reyes, A.-R.S., Mercado-Berton, A., Ong, C., Cohen, S., Hu, L., et al. (2012). Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. Nat Neurosci *15*, 274–283.

Goffin, D., Brodkin, E.S., Blendy, J.A., Siegel, S.J., and Zhou, Z. (2014). Cellular origins of auditory event-related potential deficits in Rett syndrome. Nat. Neurosci. *17*, 804–806.

Gong, S., Doughty, M., Harbaugh, C.R., Cummins, A., Hatten, M.E., Heintz, N., and Gerfen, C.R. (2007). Targeting Cre Recombinase to Specific Neuron Populations with Bacterial Artificial Chromosome Constructs. J. Neurosci. *27*, 9817–9823.

Großer, E., Hirt, U., Janc, O.A., Menzfeld, C., Fischer, M., Kempkes, B., Vogelgesang, S., Manzke, T.U., Opitz, L., Salinas-Riester, G., et al. (2012). Oxidative burden and mitochondrial dysfunction in a mouse model of Rett syndrome. Neurobiol. Dis. *48*, 102–114.

Guan, J.-S., Haggarty, S.J., Giacometti, E., Dannenberg, J.-H., Joseph, N., Gao, J., Nieland, T.J.F., Zhou, Y., Wang, X., Mazitschek, R., et al. (2009). HDAC2 negatively regulates memory formation and synaptic plasticity. Nature *459*, 55–60.

Guan, Z., Giustetto, M., Lomvardas, S., Kim, J.-H., Miniaci, M.C., Schwartz, J.H., Thanos, D., and Kandel, E.R. (2002). Integration of Long-Term-Memory-Related Synaptic Plasticity Involves Bidirectional Regulation of Gene Expression and Chromatin Structure. Cell *111*, 483–493.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. *17*, 215–222.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet *27*, 322–6.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science *315*, 1143–7.

Hagberg, B., Aicardi, J., Dias, K., and Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. Ann. Neurol. *14*, 471–479.

Hansen, J.C., Ghosh, R.P., and Woodcock, C.L. (2010). Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. IUBMB Life *62*, 732–738.

Hay, R.T. (2005). SUMO: a history of modification. Mol. Cell *18*, 1–12.

He, L., Liu, N., Cheng, T., Chen, X., Li, Y., Shu, Y., Qiu, Z., and Zhang, X. (2014). Conditional deletion of Mecp2 in parvalbumin-expressing GABAergic cells results in the absence of critical period plasticity. Nat. Commun. *5*, 5036.

Heckman, L.D., Chahrour, M.H., and Zoghbi, H.Y. (2014). Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. eLife *3*, e02676.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., nas, M.S.-F., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738–48.

Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. Mol. Cell *29*, 525–531.

Holtmaat, A., and Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. Nat. Rev. Neurosci. *10*, 647–658.

Horike, S., Cai, S., Miyano, M., Cheng, J.-F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat. Genet. *37*, 31–40.

Im, H.-I., Hollander, J.A., Bali, P., and Kenny, P.J. (2010). MeCP2 controls BDNF expression and cocaine intake through homeostatic interactions with microRNA-212. Nat. Neurosci. *13*, 1120–1127.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

Ishibashi, T., Thambirajah, A.A., and Ausió, J. (2008). MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation. FEBS Lett. *582*, 1157–1162.

Itoh, M., Tahimic, C.G.T., Ide, S., Otsuki, A., Sasaoka, T., Noguchi, S., Oshimura, M., Goto, Y., and Kurimasa, A. (2012). Methyl CpG-binding Protein Isoform MeCP2_e2 Is Dispensable for Rett Syndrome Phenotypes but Essential for Embryo Viability and Placenta Development. J. Biol. Chem. *287*, 13859–13867.

Ito-Ishida, A., Ure, K., Chen, H., Swann, J.W., and Zoghbi, H.Y. (2015). Loss of MeCP2 in Parvalbumin-and Somatostatin-Expressing Neurons in Mice Leads to Distinct Rett Syndrome-like Phenotypes. Neuron *88*, 651–658.

Jain, S., and Parker, R. (2013). The discovery and analysis of P Bodies. Adv. Exp. Med. Biol. *768*, 23–43.

Jeffery, L., and Nakielny, S. (2004). Components of the DNA Methylation System of Chromatin Control Are RNA-binding Proteins. J. Biol. Chem. *279*, 49479–49487.

Jentarra, G.M., Olfers, S.L., Rice, S.G., Srivastava, N., Homanics, G.E., Blue, M., Naidu, S., and Narayanan, V. (2010). Abnormalities of cell packing density and dendritic complexity in the MeCP2 A140V mouse model of Rett syndrome/X-linked mental retardation. BMC Neurosci. *11*, 19.

Jian, L., Nagarajan, L., de Klerk, N., Ravine, D., Bower, C., Anderson, A., Williamson, S., Christodoulou, J., and Leonard, H. (2006). Predictors of seizure onset in Rett syndrome. J. Pediatr. *149*, 542–547.

Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanović, D., Geschwind, D.H., Mane, S.M., State, M.W., and Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron *62*, 494–509.

Jordan, C., Li, H.H., Kwan, H.C., and Francke, U. (2007). Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. BMC Med. Genet. *8*, 36.

Kankirawatana, P., Leonard, H., Ellaway, C., Scurlock, J., Mansour, A., Makris, C.M., Dure, L.S., Friez, M., Lane, J., Kiraly-Borri, C., et al. (2006). Early progressive encephalopathy in boys and MECP2 mutations. Neurology *67*, 164–166.

Kann, O., Papageorgiou, I.E., and Draguhn, A. (2014). Highly Energized Inhibitory Interneurons are a Central Element for Information Processing in Cortical Networks. J. Cereb. Blood Flow Metab. *34*, 1270–1282.

Katz, D.M., Berger-Sweeney, J.E., Eubanks, J.H., Justice, M.J., Neul, J.L., Pozzo-Miller, L., Blue, M.E., Christian, D., Crawley, J.N., Giustetto, M., et al. (2012). Preclinical research in Rett syndrome: setting the foundation for translational success. Dis. Model. Mech. *5*, 733–745.

Kaufmann, W.E., Tierney, E., Rohde, C.A., Suarez-Pedraza, M.C., Clarke, M.A., Salorio, C.F., Bibat, G., Bukelis, I., Naram, D., Lanham, D.C., et al. (2012). Social impairments in Rett syndrome: characteristics and relationship with clinical severity. J. Intellect. Disabil. Res. *56*, 233–247.

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R.W., and Bérubé, N.G. (2010). ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. Dev. Cell *18*, 191–202.

Kernohan, K.D., Vernimmen, D., Gloor, G.B., and Bérubé, N.G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic Acids Res. *42*, 8356–8368.

Kerr, B., Soto C, J., Saez, M., Abrams, A., Walz, K., and Young, J.I. (2012). Transgenic complementation of MeCP2 deficiency: phenotypic rescue of Mecp2-null mice by isoform-specific transgenes. Eur. J. Hum. Genet. EJHG *20*, 69–76.

Khwaja, O.S., Ho, E., Barnes, K.V., O'Leary, H.M., Pereira, L.M., Finkelstein, Y., Nelson, C.A., Vogel-Farley, V., DeGregorio, G., Holm, I.A., et al. (2014). Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. Proc. Natl. Acad. Sci. U. S. A. *111*, 4596–4601.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nat Protoc *4*, 506–17.

Kim, K.-Y., Hysolli, E., and Park, I.-H. (2011). Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome. Proc. Natl. Acad. Sci. U. S. A. *108*, 14169–14174.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.-S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., et al. (2013). Topoisomerases facilitate transcription of long genes linked to autism. Nature *501*, 58–62.

Kishi, N., and Macklis, J.D. (2004). MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. Mol. Cell. Neurosci. *27*, 306–321.

Kishi, N., and Macklis, J.D. (2010). MeCP2 functions largely cell-autonomously, but also non-cell-autonomously, in neuronal maturation and dendritic arborization of cortical pyramidal neurons. Exp. Neurol. *222*, 51–58.

Klein, M.E., Lioy, D.T., Ma, L., Impey, S., Mandel, G., and Goodman, R.H. (2007). Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. Nat. Neurosci. *10*, 1513–1514.

Klose, R.J., and Bird, A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J. Biol. Chem. *279*, 46490–46496.

Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG. Mol. Cell *19*, 667–678.

Kokura, K., Kaul, S.C., Wadhwa, R., Nomura, T., Khan, M.M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. J. Biol. Chem. *276*, 34115–34121.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat. Struct. Mol. Biol. *20*, 396–403.

Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res. *32*, 1818–1823.

Kriaucionis, S., Paterson, A., Curtis, J., Guy, J., MacLeod, N., and Bird, A. (2006). Gene Expression Analysis Exposes Mitochondrial Abnormalities in a Mouse Model of Rett Syndrome. Mol. Cell. Biol. *26*, 5033–5042.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Dragich, J., Schanen, C., and Tamura, M. (2001). Functional analyses of MeCP2 mutations associated with Rett syndrome using transient expression systems. Brain Dev. *23*, *Supplement 1*, S165–S173.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Hammer, S., Schanen, C., Terai, I., and Tamura, M. (2002). Functional characterisation of MeCP2 mutations found in male patients with X linked mental retardation. J. Med. Genet. *39*, 132–136.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C., and Tamura, M. (2003). Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. J. Med. Genet. *40*, 487–493.

Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M., and Schanen, N.C. (2008a). Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. J. Cell Sci. *121*, 1128–1137.

Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M., and Schanen, N.C. (2008b). Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. J. Cell Sci. *121*, 1128–1137.

Landi, S., Putignano, E., Boggio, E.M., Giustetto, M., Pizzorusso, T., and Ratto, G.M. (2011). The short-time structural plasticity of dendritic spines is altered in a model of Rett syndrome. Sci. Rep. *1*, 45.

Lawson-Yuen, A., Liu, D., Han, L., Jiang, Z.I., Tsai, G.E., Basu, A.C., Picker, J., Feng, J., and Coyle, J.T. (2007). Ube3a mRNA and protein expression are not decreased in Mecp2R168X mutant mice. Brain Res. *1180*, 1–6.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. Cell *69*, 905–914.

Li, H., Zhong, X., Chau, K.F., Williams, E.C., and Chang, Q. (2011). Loss of activity-induced phosphorylation of MeCP2 enhances synaptogenesis, LTP and spatial memory. Nat. Neurosci. *14*, 1001–1008.

Li, Y., Wang, H., Muffat, J., Cheng, A.W., Orlando, D.A., Lovén, J., Kwok, S., Feldman, D.A., Bateup, H.S., Gao, Q., et al. (2013a). Global Transcriptional and Translational Repression in Human-Embryonic-Stem-Cell-Derived Rett Syndrome Neurons. Cell Stem Cell *13*, 446–458.

Li, Y.R., King, O.D., Shorter, J., and Gitler, A.D. (2013b). Stress granules as crucibles of ALS pathogenesis. J. Cell Biol. *201*, 361–372.

Linhoff, M.W., Garg, S.K., and Mandel, G. (2015). A High-Resolution Imaging Approach to Investigate Chromatin Architecture in Complex Tissues. Cell *163*, 246–255.

Lioy, D.T., Garg, S.K., Monaghan, C.E., Raber, J., Foust, K.D., Kaspar, B.K., Hirrlinger, P.G., Kirchhoff, F., Bissonnette, J.M., Ballas, N., et al. (2011). A role for glia in the progression of Rett/'s syndrome. Nature *advance online publication*.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. Science *341*, 1237905.

Liu, Z., Li, X., Zhang, J.-T., Cai, Y.-J., Cheng, T.-L., Cheng, C., Wang, Y., Zhang, C.-C., Nie, Y.-H., Chen, Z.-F., et al. (2016). Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. Nature *530*, 98–102.

Lombardi, L.M., Baker, S.A., and Zoghbi, H.Y. (2015). MECP2 disorders: from the clinic to mice and back. J. Clin. Invest. *125*, 2914–2923.

Luikenhuis, S., Giacometti, E., Beard, C.F., and Jaenisch, R. (2004). Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. Proc. Natl. Acad. Sci. U. S. A. *101*, 6033–6038.

Lyst, M.J., and Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nat. Rev. Genet. *16*, 261–275.

Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci. *16*, 898–902.

Mabb, A.M., Kullmann, P.H.M., Twomey, M.A., Miriyala, J., Philpot, B.D., and Zylka, M.J. (2014). Topoisomerase 1 inhibition reversibly impairs synaptic function. Proc. Natl. Acad. Sci. *111*, 17290–17295.

Madabhushi, R., Gao, F., Pfenning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.-C., et al. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell *161*, 1592–1605.

Maffei, A., and Fontanini, A. (2009). Network homeostasis: a matter of coordination. Curr. Opin. Neurobiol. *19*, 168–173.

Malik, H.S., and Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. Cell *18*, 1030–1040.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells. Cell *143*, 527–539.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. Cell *161*, 541–554.

McGraw, C.M., Samaco, R.C., and Zoghbi, H.Y. (2011). Adult Neural Function Requires MeCP2. Science *333*, 186–186.

Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K., and Epplen, J.T. (2005). Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. J. Med. Genet. *42*, e12–e12.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell *151*, 1417–1430.

Meloni, I., Bruttini, M., Longo, I., Mari, F., Rizzolio, F., D'Adamo, P., Denvriendt, K., Fryns, J.-P., Toniolo, D., and Renieri, A. (2000). A Mutation in the Rett Syndrome Gene, MECP2, Causes X-Linked Mental Retardation and Progressive Spasticity in Males. Am. J. Hum. Genet. *67*, 982–985.

Métivier, R., Penot, G., Hübner, M.R., Reid, G., Brand, H., Koš, M., and Gannon, F. (2003). Estrogen Receptor-α Directs Ordered, Cyclical, and Combinatorial Recruitment of Cofactors on a Natural Target Promoter. Cell *115*, 751–763.

Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J.M., Jones, J.R., Scherer, S.W., Schanen, N.C., Friez, M.J., et al. (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat. Genet. *36*, 339–341.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R., et al. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369–1384.

Monory, K., Massa, F., Egertová, M., Eder, M., Blaudzun, H., Westenbroek, R., Kelsch, W., Jacob, W., Marsch, R., Ekker, M., et al. (2006). The Endocannabinoid System Controls Key Epileptogenic Circuits in the Hippocampus. Neuron *51*, 455–466.

Mullaney, B.C., Johnston, M.V., and Blue, M.E. (2004). Developmental expression of methyl-CpG binding protein 2 is dynamically regulated in the rodent brain. Neuroscience *123*, 939–49.

Müller, M., and Can, K. (2014). Aberrant redox homoeostasis and mitochondrial dysfunction in Rett syndrome. Biochem. Soc. Trans. *42*, 959–964.

Muotri, A.R., Marchetto, M.C.N., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443–446.

Murakami, J.W., Courchesne, E., Haas, R.H., Press, G.A., and Yeung-Courchesne, R. (1992). Cerebellar and cerebral abnormalities in Rett syndrome: a quantitative MR analysis. AJR Am. J. Roentgenol. *159*, 177–183.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Nagano, T., Lubling, Y., Yaffe, E., Wingett, S.W., Dean, W., Tanay, A., and Fraser, P. (2015). Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. Nat. Protoc. *10*, 1986–2003.

Nan, X., Tate, P., Li, E., and Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. Mol. Cell. Biol. *16*, 414–421.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471–481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–9.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc. Natl. Acad. Sci. *104*, 2709–2714.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat. Struct. Mol. Biol. *20*, 387–395.

Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E.S., Schanen, N.C., Zappella, M., et al. (2010). Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. Ann. Neurol. *68*, 944–950.

Nguyen, M.V.C., Du, F., Felice, C.A., Shan, X., Nigam, A., Mandel, G., Robinson, J.K., and Ballas, N. (2012). MeCP2 is critical for maintaining mature neuronal networks and global brain anatomy during late stages of postnatal brain development and in the mature adult brain. J. Neurosci. Off. J. Soc. Neurosci. *32*, 10021–10034.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007a). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Mol. Cell. Biol. *27*, 864–77.

Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A., and Woodcock, C.L. (2007b). MeCP2-Chromatin Interactions Include the Formation of Chromatosome-like Structures and Are Altered in Mutations Causing Rett Syndrome. J. Biol. Chem. *282*, 28237–28245.

Nishiyama, M., Skoultchi, A.I., and Nakayama, K.I. (2012). Histone H1 Recruitment by CHD8 Is Essential for Suppression of the Wnt–β-Catenin Signaling Pathway. Mol. Cell. Biol. *32*, 501–512.

Nuber, U.A., Kriaucionis, S., Roloff, T.C., Guy, J., Selfridge, J., Steinhoff, C., Schulz, R., Lipkowitz, B., Ropers, H.H., Holmes, M.C., et al. (2005). Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. Hum. Mol. Genet. *14*, 2247–2256.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. Nat. Neurosci. *11*, 1271–1282.

O'Leary, T., Williams, A.H., Franci, A., and Marder, E. (2014). Cell Types, Network Homeostasis, and Pathological Compensation from a Biologically Plausible Ion Channel Expression Model. Neuron *82*, 809–821.

Ooi, S.L., Henikoff, J.G., and Henikoff, S. (2010). A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. Nucleic Acids Res *38*, e26.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. *36*, 1065–1071.

Padmanabhan, K., and Urban, N.N. (2010). Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. Nat. Neurosci. *13*, 1276–1282.

Pelka, G.J., Watson, C.M., Christodoulou, J., and Tam, P.P.L. (2005). Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. Genomics *85*, 441–452.

Pelka, G.J., Watson, C.M., Radziewic, T., Hayward, M., Lahooti, H., Christodoulou, J., and Tam, P.P.L. (2006). Mecp2 deficiency is associated with learning and cognitive deficits and altered gene activity in the hippocampal region of mice. Brain *129*, 887–898.

Percy, A.K., Neul, J.L., Glaze, D.G., Motil, K.J., Skinner, S.A., Khwaja, O., Lee, H.-S., Lane, J.B., Barrish, J.O., Annese, F., et al. (2010). Rett syndrome diagnostic criteria: Lessons from the Natural History Study. Ann. Neurol. *68*, 951–955.

Perissi, V., Jepsen, K., Glass, C.K., and Rosenfeld, M.G. (2010). Deconstructing repression: evolving models of co-repressor action. Nat. Rev. Genet. *11*, 109–123.

Pitcher, M.R., Ward, C.S., Arvide, E.M., Chapleau, C.A., Pozzo-Miller, L., Hoeflich, A., Sivaramakrishnan, M., Saenger, S., Metzger, F., and Neul, J.L. (2013). Insulinotropic treatments exacerbate metabolic syndrome in mice lacking MeCP2 function. Hum. Mol. Genet. *22*, 2626–2633.

Pitcher, M.R., Herrera, J.A., Buffington, S.A., Kochukov, M.Y., Merritt, J.K., Fisher, A.R., Schanen, N.C., Costa-Mattioli, M., and Neul, J.L. (2015). Rett syndrome like phenotypes in the R255X Mecp2 mutant mouse are rescued by MECP2 transgene. Hum. Mol. Genet. ddv030.

Potter, G.B., Petryniak, M.A., Shevchenko, E., McKinsey, G.L., Ekker, M., and Rubenstein, J.L.R. (2009). Generation of Cre-transgenic mice using Dlx1/Dlx2 enhancers and their characterization in GABAergic interneurons. Mol. Cell. Neurosci. *40*, 167–186.

Puc, J., Kozbial, P., Li, W., Tan, Y., Liu, Z., Suter, T., Ohgi, K.A., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2015). Ligand-Dependent Enhancer Activation Regulated by Topoisomerase-I Activity. Cell *160*, 367–380.

Qiu, Z., Sylwestrak, E.L., Lieberman, D.N., Zhang, Y., Liu, X.-Y., and Ghosh, A. (2012). The Rett Syndrome Protein MeCP2 Regulates Synaptic Scaling. J. Neurosci. *32*, 989–994.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216–226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLOS Biol *4*, e309.

Ramirez, J.-M., Ward, C.S., and Neul, J.L. (2013). Breathing challenges in Rett syndrome: lessons learned from humans and animal models. Respir. Physiol. Neurobiol. *189*, 280–287.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Ravn, K., Nielsen, J.B., Uldall, P., Hansen, F.J., and Schwartz, M. (2003). No correlation between phenotype and genotype in boys with a truncating MECP2 mutation. J. Med. Genet. *40*, e5.

Reichwald, K., Thiesen, J., Wiehe, T., Weitzel, J., Strätling, W.H., Kioschis, P., Poustka, A., Rosenthal, A., and Platzer, M. (2014). Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. Mamm. Genome *11*, 182–190.

Reiss, A.L., Faruque, F., Naidu, S., Abrams, M., Beaty, T., Bryan, R.N., and Moser, H. (1993). Neuroanatomy of Rett syndrome: A volumetric imaging study. Ann. Neurol. *34*, 227–234.

Renieri, A., Mari, F., Mencarelli, M.A., Scala, E., Ariani, F., Longo, I., Meloni, I., Cevenini, G., Pini, G., Hayek, G., et al. (2009). Diagnostic criteria for the Zappella variant of Rett syndrome (the preserved speech variant). Brain Dev. *31*, 208–216.

Rett, A. (1966). [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. Wien. Med. Wochenschr. 1946 *116*, 723–726.

Rube, H.T., Lee, W., Hejna, M., Chen, H., Yasui, D.H., Hess, J.F., LaSalle, J.M., Song, J.S., and Gong, Q. (2016). Sequence features accurately predict genome-wide MeCP2 binding in vivo. Nat. Commun. *7*, 11025.

Samaco, R.C., Fryer, J.D., Ren, J., Fyffe, S., Chao, H.-T., Sun, Y., Greer, J.J., Zoghbi, H.Y., and Neul, J.L. (2008). A partial loss of function allele of Methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. Hum. Mol. Genet. *17*, 1718–1727.

Samaco, R.C., Mandel-Brehm, C., Chao, H.-T., Ward, C.S., Fyffe-Maricich, S.L., Ren, J., Hyland, K., Thaller, C., Maricich, S.M., Humphreys, P., et al. (2009). Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis and specific behavioral abnormalities. Proc. Natl. Acad. Sci. *106*, 21966–21971.

Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012). Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. Nat. Genet. *44*, 206–211.

Saunders, C.J., Minassian, B.E., Chow, E.W.C., Zhao, W., and Vincent, J.B. (2009). Novel exon 1 mutations in MECP2 implicate isoform MeCP2_e1 in classical Rett syndrome. Am. J. Med. Genet. A. *149A*, 1019–1023.

Saxena, A., de Lagarde, D., Leonard, H., Williamson, S.L., Vasudevan, V., Christodoulou, J., Thompson, E., MacLeod, P., and Ravine, D. (2006). Lost in translation: translational interference from a recurrent mutation in exon 1 of MECP2. J. Med. Genet. *43*, 470–477.

Schmiedeberg, L., Skene, P., Deaton, A., and Bird, A. (2009). A Temporal Threshold for Formaldehyde Crosslinking and Fixation. PLoS ONE *4*, e4636.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. *18*, 1251–1262.

Seok, J., Xu, W., Gao, H., Davis, R.W., and Xiao, W. (2012). JETTA: junction and exon toolkits for transcriptome analysis. Bioinformatics *28*, 1274–1275.

Sephton, C.F., Cenik, C., Kucukural, A., Dammer, E.B., Cenik, B., Han, Y., Dewey, C.M., Roth, F.P., Herz, J., Peng, J., et al. (2011). Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. J. Biol. Chem. *286*, 1204–1215.

Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002a). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron *35*, 243–54.

Shahbazian, M.D., Antalffy, B., Armstrong, D.L., and Zoghbi, H.Y. (2002b). Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum. Mol. Genet. *11*, 115–124.

Singh, J., Saxena, A., Christodoulou, J., and Ravine, D. (2008). MECP2 genomic structure and function: insights from ENCODE. Nucleic Acids Res. *36*, 6035–6047.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. Mol. Cell *37*, 457–468.

Subramaniam, B., Naidu, S., and Reiss, A.L. (1997). Neuroanatomy in Rett syndrome Cerebral cortex and posterior fossa. Neurology *48*, 399–407.

Sugino, K., Hempel, C.M., Okaty, B.W., Arnson, H.A., Kato, S., Dani, V.S., and Nelson, S.B. (2014). Cell-Type-Specific Repression by Methyl-CpG-Binding Protein 2 Is Biased toward Long Genes. J. Neurosci. *34*, 12877–12883.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci *14*, 1607–1616.

Tai, D.J.C., Liu, Y.C., Hsu, W.L., Ma, Y.L., Cheng, S.J., Liu, S.Y., and Lee, E.H.Y. (2016). MeCP2 SUMOylation rescues Mecp2-mutant-induced behavioural deficits in a mouse model of Rett syndrome. Nat. Commun. *7*, 10552.

Takeshima, H., Suetake, I., and Tajima, S. (2008). Mouse Dnmt3a preferentially methylates linker DNA and is inhibited by histone H1. J. Mol. Biol. *383*, 810–821.

Tao, J., Hu, K., Chang, Q., Wu, H., Sherman, N.E., Martinowich, K., Klose, R.J., Schanen, C., Jaenisch, R., Wang, W., et al. (2009). Phosphorylation of MeCP2 at Serine 80 regulates its chromatin association and neurological function. Proc. Natl. Acad. Sci. U. S. A. *106*, 4882–4887.

Tarquinio, D.C., Motil, K.J., Hou, W., Lee, H.-S., Glaze, D.G., Skinner, S.A., Neul, J.L., Annese, F., McNair, L., Barrish, J.O., et al. (2012). Growth failure and outcome in Rett syndrome. Neurology *79*, 1653–1661.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.

Temudo, T., Santos, M., Ramos, E., Dias, K., Vieira, J.P., Moreira, A., Calado, E., Carrilho, I., Oliveira, G., Levy, A., et al. (2011). Rett syndrome with and without detected MECP2 mutations: An attempt to redefine phenotypes. Brain Dev. *33*, 69–76.

Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. Nat. Struct. Mol. Biol. *21*, 88–94.

Thambirajah, A.A., Ng, M.K., Frehlick, L.J., Li, A., Serpa, J.J., Petrotchenko, E.V., Silva-Moreno, B., Missiaen, K.K., Borchers, C.H., Adam Hall, J., et al. (2012). MeCP2 binds to nucleosome free (linker DNA) regions and to H3K9/H3K27 methylated nucleosomes in the brain. Nucleic Acids Res. *40*, 2884–2897.

Trask, H.W., Cowper-Sal-lari, R., Sartor, M.A., Gui, J., Heath, C.V., Renuka, J., Higgins, A.-J., Andrews, P., Korc, M., Moore, J.H., et al. (2009). Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. RNA *15*, 1917–1928.

Traynor, J., Agarwal, P., Lazzeroni, L., and Francke, U. (2002). Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. BMC Med. Genet. *3*, 12.

Tsao, Y.-P., Wu, H.-Y., and Liu, L.F. (1989). Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. Cell *56*, 111–118.

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc. Natl. Acad. Sci. *99*, 15536–15541.

Turrigiano, G. (2011). Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement. Annu. Rev. Neurosci. *34*, 89–103.

Urdinguio, R.G., Lopez-Serra, L., Lopez-Nieva, P., Alaminos, M., Diaz-Uriarte, R., Fernandez, A.F., and Esteller, M. (2008). Mecp2 -Null Mice Provide New Neuronal Targets for Rett Syndrome. PLOS ONE *3*, e3669.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.

Valenti, D., de Bari, L., De Filippis, B., Henrion-Caude, A., and Vacca, R.A. (2014). Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: An overview of Down syndrome, autism, Fragile X and Rett syndrome. Neurosci. Biobehav. Rev. *46*, *Part 2*, 202–217.

Venâncio, M., Santos, M., Pereira, S.A., Maciel, P., and Saraiva, J.M. (2007). An explanation for another familial case of Rett syndrome: maternal germline mosaicism. Eur. J. Hum. Genet. EJHG *15*, 902–904.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat. Rev. Mol. Cell Biol. *3*, 430–440.

Wang, I.-T.J., Reyes, A.-R.S., and Zhou, Z. (2013). Neuronal morphology in MeCP2 mouse models is intrinsically variable and depends on age, cell type, and Mecp2 mutation. Neurobiol. Dis. *58*, 3–12.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897–903.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell *138*, 1019–1031.

Ward, C.S., Arvide, E.M., Huang, T.-W., Yoo, J., Noebels, J.L., and Neul, J.L. (2011). MeCP2 Is Critical within HoxB1-Derived Tissues of Mice for Normal Lifespan. J. Neurosci. *31*, 10359–10370.

Weaving, L.S., Williamson, S.L., Bennetts, B., Davis, M., Ellaway, C.J., Leonard, H., Thong, M.-K., Delatycki, M., Thompson, E.M., Laing, N., et al. (2003). Effects of MECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. Am. J. Med. Genet. A. *118A*, 103–114.

Wegener, E., Brendel, C., Fischer, A., Hülsmann, S., Gärtner, J., and Huppke, P. (2014). Characterization of the MeCP2 R168X Knockin Mouse Model for Rett Syndrome. PLOS ONE *9*, e115444.

Werner, M.S., and Ruthenburg, A.J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Rep. *12*, 1089–1098.

Wood, L., Gray, N.W., Zhou, Z., Greenberg, M.E., and Shepherd, G.M.G. (2009). Synaptic Circuit Abnormalities of Motor-Frontal Layer 2/3 Pyramidal Neurons in an RNA Interference Model of Methyl-CpG-Binding Protein 2 Deficiency. J. Neurosci. *29*, 12440–12448.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science *329*, 444–448.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., Smallwood, P.M., Erlanger, B., Wheelan, S.J., and Nathans, J. (2014). Cellular Resolution Maps of X Chromosome Inactivation: Implications for Neural Development, Function, and Disease. Neuron *81*, 103–119.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell *154*, 1370–1379.

Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc. Natl. Acad. Sci. U. S. A. *104*, 19416–19421.

Yazdani, M., Deogracias, R., Guy, J., Poot, R.A., Bird, A., and Barde, Y.-A. (2012). Disease Modeling Using Embryonic Stem Cells: MeCP2 Regulates Nuclear Size and RNA Synthesis in Neurons. STEM CELLS *30*, 2128–2139.

Young, D., Bebbington, A., de Klerk, N., Bower, C., Nagarajan, L., and Leonard, H. (2011). The relationship between MECP2 mutation type and health status and service use trajectories over time in a Rett syndrome population. Res. Autism Spectr. Disord. *5*, 442–449.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., et al. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proc. Natl. Acad. Sci. U. S. A. *102*, 17551–17558.

Yu, F., Thiesen, J., and Strätling, W.H. (2000). Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. Nucleic Acids Res. *28*, 2201–2206.

Zachariah, R.M., Olson, C.O., Ezeonwuka, C., and Rastegar, M. (2012). Novel MeCP2 Isoform-Specific Antibody Reveals the Endogenous MeCP2E1 Expression in Murine Brain, Primary Neurons and Astrocytes. PloS One *7*, e49763.

Zappella, M., Meloni, I., Longo, I., Canitano, R., Hayek, G., Rosaia, L., Mari, F., and Renieri, A. (2003). Study of MECP2 gene in Rett syndrome variants and autistic girls. Am. J. Med. Genet. B Neuropsychiatr. Genet. *119B*, 102–107.

Zeev, B.B., Yaron, Y., Schanen, N.C., Wolf, H., Brandt, N., Ginot, N., Shomrat, R., and Orr-Urtreger, A. (2002). Rett Syndrome: Clinical Manifestations in Males With MECP2 Mutations. J. Child Neurol. *17*, 20–24.

Zentner, G.E., Tsukiyama, T., and Henikoff, S. (2013). ISWI and CHD Chromatin Remodelers Bind Promoters but Act in Gene Bodies. PLOS Genet *9*, e1003317.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17.

Zhong, X., Li, H., and Chang, Q. (2012). MeCP2 Phosphorylation Is Required for Modulating Synaptic Scaling through mGluR5. J. Neurosci. *32*, 12841–12847.

Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. Neuron *52*, 255–69.

Zovkic, I.B., Guzman-Karlsson, M.C., and Sweatt, J.D. (2013). Epigenetic regulation of memory formation and maintenance. Learn. Mem. *20*, 61–74.

Zylka, M.J., Simon, J.M., and Philpot, B.D. (2015). Gene Length Matters in Neurons. Neuron *86*, 353–355.